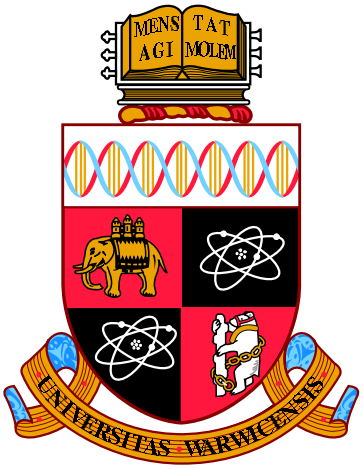


Engineering of RNA sensors and actuators in living cells

by

William Galdric Pierre Maddison Rostain

Thesis

Submitted for the degree of

Doctor of Philosophy

School of Life Sciences

The University of Warwick

June 2017

Une science, au moment où elle naît, non seulement se débarrasse d'un certain nombre d'obstacles, mais en même temps, elle supprime un certain nombre de savoirs et de connaissances existantes qu'elle occulte, qu'elle cache, comme si elle appliquait une grille nouvelle qui tout en permettant de faire apparaître des phénomènes jusque là cachés, cache des connaissances déjà acquises. Donc une science, ce n'est pas simplement l'oubli des vieux préjugés, c'est une véritable nouvelle grille qui cache et qui du coup fait apparaître.

— Michel Foucault

To my glipglops…

Contents

List of Figures

List of Tables

List of abbreviations

2D-PAGE : 2 Dimensional Polyacrylamide Gel Electrophoresis

asRNA: Antisense RNA

aTc: Anhydrotetracycline

ATP, dATP : Adenosine Triphosphate, deoxy-ATP

bp : Base pairs

cat : Chloramphenicol Acetyl-Transferase

circRNA : Circular RNA

Cmk / *cmk*: Cytidine Monophosphate Kinase

CDP, dCDP: Cytidine Diphosphate, deoxy-CDP

CMP, dCMP: Cytidine Monophosphate, deoxy-CMP

CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats

CRISPRa : CRISPR activation

CRISPRi : CRISPR interference

cr-mRNA : Cis-repressed mRNA (sometimes referred to as crRNA in the literature)

crRNA : CRISPR RNA

DMSO : Dimethyl Sulfoxide

List of Tables

dsDNA : Double stranded DNA / RNA

dNTP : deoxy-Nucleotide Triphosphate

DTT : Dithiothreitol

EB : Elution buffer

E. coli : *Escherichia coli*

EDTA : Ethylenediaminetetraacetic acid

FACS : Fluorescence-Activated Cell Sorting

GFP : Green Fluorescent Protein

gRNA : [CRISPR] guide RNA

GTP, dGTP : Guanosine Triphosphate, deoxy-GTP

HDV : Hepatitis D virus

HHR : Hammerhead ribozyme

HR : Homologous recombination

IPTG : β-D-1-thiogalactopyranoside

ISO buffer : Isothermal amplification buffer

kbp : Kilo Base Pairs, Thousands of base pairs

LB : Lysogeny broth

MCP : MS2 coat protein

MFE / mfe : Minimum Free Energy

miRNA : micro RNA

mRNA : messenger RNA

NAD : Nicotinamide adenine dinucleotide

NEB : New England Biolabs

PACE : Phage Assisted Continuous Evolution

PAGE : Polyacrylamide Gel Electrophoresis

PAM : Protospacer-adjacent motif, or Polyacrylamide

PCP : PP7 coat protein

PCR : Polymerase Chain Reaction

PEG : Polyethylene glycol

pfu : Plaque Forming Unit

PIE ribozyme : Permuted Intron Exon ribozyme

RBS: Ribosome Binding Site

Riboregulation : RNA based regulation

RNA : Ribonucleic acid

RNase : Ribonuclease

RT : Reverse Transcription / Reverse Transcriptase

RT-PCR : Reverse Transcriptase PCR

S. pyogenes : *Streptococcus pyogenes*

SELEX : Systematic Evolution of Ligands by EXponential enrichment

sfGFP : Superfolder Green Fluorescent Protein

sgRNA : [CRISPR] synthetic guide RNA

siRNA : Small interfering RNA

sRNA : Small RNA

ssDNA / ssDNA : Single stranded DNA / RNA

snoRNA : Small nucleolar RNA

taRNA : Trans-activating RNA

TE buffer: Tris EDTA buffer

tracrRNA : Trans Activating crRNA

tRNA : Transfer RNA

TTP, dTTP : Thymidine Triphosphate, deoxy-TTP

UTR : Untranslated Region (of mRNA)

WT : Wild type

YAC : Yeast Artificial Chromosome

YUNR : pYrimidine-Uracil-Nucleotide-puRine [motif]

Acknowledgements

First, I would like to express my gratitude to Professor Alfonso Jaramillo for his support of my Ph.D study and research, as well as to all the other people I have worked with over these four years in Evry and in Warwick, for their help and support.

I would also like to thank the DGA, the Dstl and the European Commission who funded this work, and to the support staff at the University of Warwick and in the iSSB in Evry, who put up with my requests for help. I especially thank Sylvie and Bernadette, who essentially keep the iSSB going.

I thank my fellow labmates in for the stimulating discussions, nice cakes, help with experiments and motivation, and for their inextinguishable thirst for knowledge and beer. I also thank my friends in Evry, especially Boris and Costas, and my friends in Warwick, especially Despina who has very nice hair, as well as Satya, my senior colleague, Fabio, who will travel the world fixing things, Manish and John for reading over this thesis, and since I may as well name all the great people I worked with, Jack, Paul, Teresa, Eduardo, Aurelija, Michał, Rui, Thomas and the others I forgot.

Last but not least, I would like to thank my parents Bruno and Janice and my brothers Pierre-Stuart and Louis for being a good family during these four years (and also before that).

June 2017.. William Rostain

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below. The data in figure 4.6B and D were obtained by Dr. John Duncan, who also constructed the plasmids dCas9-trigger3A, dCas9-mKate2 and dCas9-mKate2ΔRBS. The dCas9 plasmid was obtained from D. Bikard (Institut Pasteur, Paris). Additionally, the PIE ribozyme sequence was sent by So Umekage (Toyohashi Uni.). The goldengate plasmids described in appendix A.1 and cisRAJ31-*cat* were built with help from David Christiany, the picture of the plate in figure 3.7 was taken by Laurianne Combes. The bioreactors used in chapter 3 were built by Fabio Polesel (U. of Warwick), who assisted me with bioreactor operation during experiments. The research was funded by a DGA/Dstl UK-France PhD programme and by E.C. grants FP7-ICT-FET 610730 (EVOPROG) and FP7-KBBE 613745 (PROMYS). All sources of information have been acknowledged by means of a reference.

Abstract

The aim of synthetic biology is to create a new discipline of engineering based on biological parts, devices and systems. The availability of predictable, programmable tools to sense and to control gene expression is central to our ability to engineer such systems. Ribonucleic acid (RNA) is an attractive building material to create such programmable tools, as RNA-RNA interactions are predictable and RNA secondary structure prediction software has been developed. Design rules for creating such parts using RNA can be established, based on a standardised approach or on structural design rules into which function is implicitly encoded. In this latter case, RNA folding software can be used to create RNA sequence which satisfy generalisable structural characteristics, but are tailored to a specific application. In this work, new design rules for the creation of RNA-based sensors and actuators are developed. The actuator parts are based on riboregulators, but with a circular topology generated through splicing of a ribozyme. The ability of these circular riboregulators to activate transcription of gene expression in *E. coli* cells is demonstrated. A method for improving these actuators by directed evolution is then tested. Finally, design rules for creating sensors of RNAs based Clustered Regularly Interspaced Short Palindromic Repeat guide RNAs (CRISPR gRNAs) are developed. These gRNA-based sensors can switch states and repress gene expression through a CRISPR-Cas9 based platform, but only in the presence of an arbitrary "trigger" RNA. The rules developed for creating sensors and actuators are characterised in *E. coli*, but are based on general principles that could be used in other organisms including eukaryotic cells.

Key words: RNA synthetic biology, computational design, riboregulator, circRNA, RNA sensing, RNA circuits, gRNA engineering, directed evolution

1 Introduction

The construction of synthetic biological circuits that can sense and interact with their environment requires the availability of sensor and actuator parts. Sensors are parts capable of sensing an input signal (e.g., the presence of a specific small molecule, or of a gene of interest), and transforming this signal into a control signal. Actuators are parts that can transform a control signal into an output, such as the expression of a reporter protein (Wang et al., 2013). In this context, a control signal usually means controlled transcription or translation of other genes.

The use of this engineering vocabulary in synthetic biology derives from a desire to encourage – and to an extent, constrain thought into – an engineering approach to the construction of genetic circuits, including the use of concepts such as standardisation and abstraction hierarchy (Endy, 2005). Applying concepts from electric and software engineering to construction of biological circuits is not straightforward, and the creation of this new engineering discipline requires the invention of new concepts and methods. To this end, working with ribonucleic acid (RNA) parts has some advantages that make it a promising building material. One of the main reasons for this is the predictable nature of RNA-RNA interactions, compared to interactions involving proteins (Kushwaha et al., 2016). As a result of our understanding of these biochemical interactions, the equilibrium structures of RNAs in solution can be computationally predicted (Zuker and Stiegler, 1981; Dirks and Pierce, 2003).

If the structure and function of an RNA are directly linked, computational prediction can generate desired function by creating RNA parts with a defined structure. When two RNA species can interact with one another, such that they have different minimum free energy (MFE) structures when folded alone or in the complexed state, and if their functionality is dependent on one of these two alternative structures, then RNA circuits that require both components to be active can be designed (Rodrigo et al., 2012). If both RNAs of the circuit are synthetic parts, their sequence can be chosen so they adopt the desired target conformations. Such parts can also be designed to interact with natural RNAs (Green et al., 2014), but the sequence choice of the synthetic RNAs are then constrained by the structure and sequence of the natural RNA. In this case, a robust and flexible design architecture is needed, so that the part can interact with a large variety of sequences. Regardless of the circuit and its use, this approach for designing RNA parts requires one or more target structures, into which the function of the RNA is implicitly encoded. This design goal is a set of sequence and structural constraints chosen such that, ideally, any RNA that meets these constraints has the desired function. The second step is the generation of the specific RNAs that meet the general constraints. Such RNAs could be designed by hand if the constraints are simple, but creating RNAs that respect a complex set of constraints rapidly becomes intractable without computationally assisted design methods. The biological context in which the RNA parts are used must also be taken into account, as other elements such as RNA stability can also affect the behaviour of the designed circuits.

Chapter 1 introduces the basic properties of RNA, RNA structure, and the methods and software used for computational prediction of RNA secondary structure. An overview of RNA parts used in the literature and design rules used to create them are then introduced, with a focus on parts that make use of RNA folding programs. Chapters 2, 3 and 4 contain more specific introductory material relevant to the strategies used to develop RNA sensors and actuators.

1.0.1 RNA structure

RNA is a polymeric molecule made of a chain of nucleotides linked through a ribose-phosphate backbone. An RNA sequence is made up of four nitrogenous bases: The purines adenine (A) and guanine (G); and the pyrimidines cytosine (C) and uracil (U). Each base is attached to the 1' carbon of a ribose sugar. A phosphodiester bond links the 3' carbon of each ribose to the 5' carbon of the next one, forming a polymer of bases on a ribose-phosphate backbone (see fig. 1.1A and B). RNA can either be single stranded, or like DNA, form a stable double helix through base pairing with a sequence containing complementary bases, where A pairs with U and G pairs with C. Other interactions that do not follow Watson-Crick base pairing rules, notably G-U base pairs (also called wobble base pairs) are also possible. G-U base pairs are relatively common in RNA structures, and for the purposes of structure prediction, are considered as normal base pairs (Varani and McClain, 2000). A-G or A-C base pairs also exist, but are unusual and are not considered by structure prediction software. Double stranded helices formed by base pairing, as well as a relatively small number of structural motifs, compose the majority of secondary structures found in RNA (Serra and Turner, 1995, see fig. 1.1C). RNA folding, like protein folding, can be broken down into primary, secondary and tertiary structure. The primary structure consists of the RNA sequence. The secondary structure is determined by the RNA folding onto itself through base pairing. Its tertiary structure consists of the higher level folding of these structures onto themselves, through chemical interactions involving the ribose-phosphate backbone, loops or bulges.

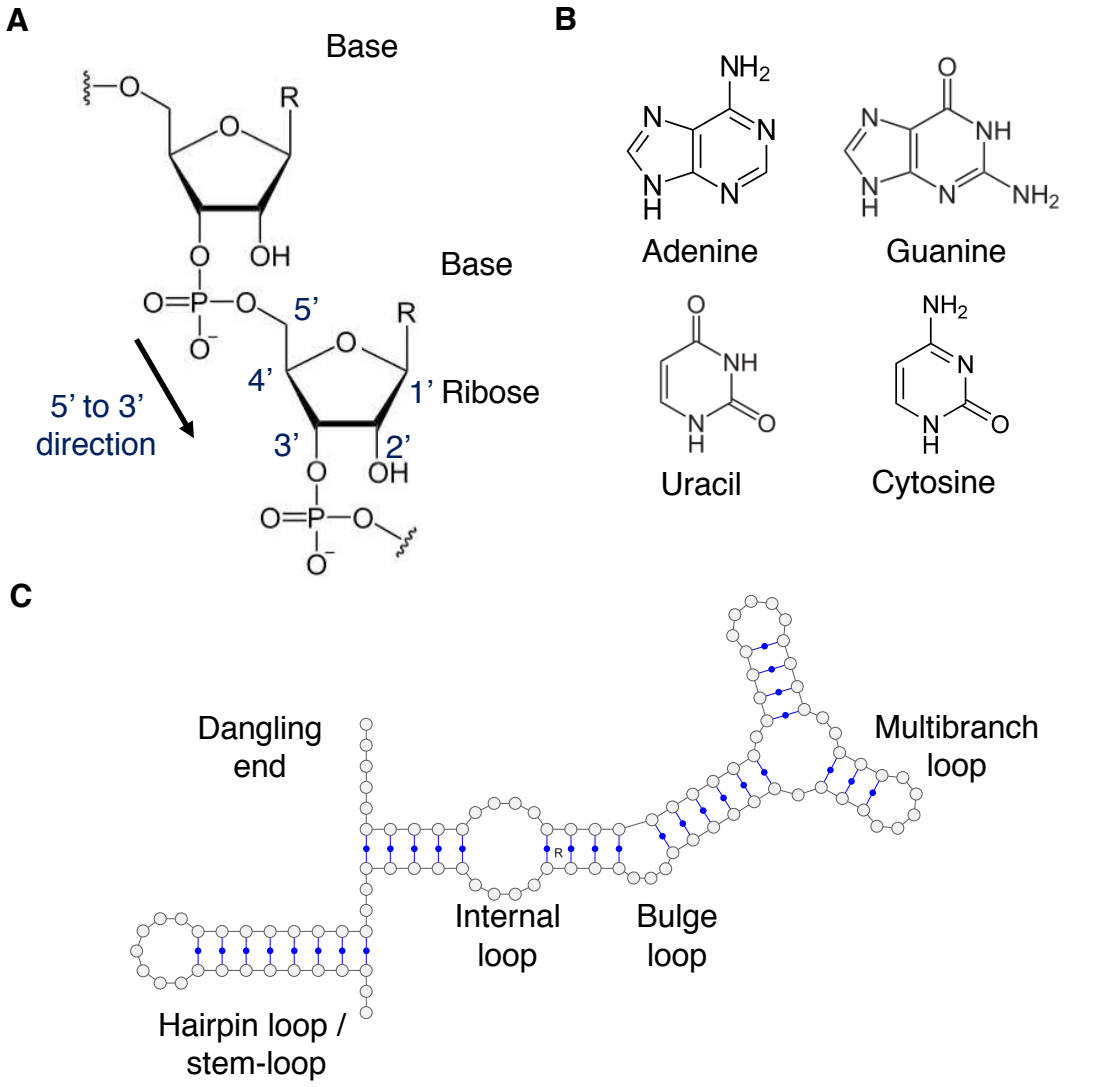


Figure 1.1 – **A**. Chemical structure of the RNA backbone. **B**. Chemical structure of the purines Adenine and Guanine, and of the pyrimidines Uracil and Cytosine. **C**. Common 2D motifs seen in RNA structures. Chemical structures adapted from Wikimedia Commons, 2D motifs adapted from Serra and Turner (1995).

1.0.2 Predicting stability of a given structure

RNA has a relatively small molecular "vocabulary". Secondary structure is the main contributor to the free energy of an RNA structure, and base pairing effects are more important than tertiary interactions for RNA folding (Banerjee et al., 1993; Laing and Draper, 1994; Mathews et al., 1997). Aside from the identity of the base pairs, the most important contributors to the stability of an RNA structure are the stacking effects of adjacent base pairs, and loop entropies (Hofacker and Stadler, 2008). A nearest neighbour model, where the thermodynamic properties of a base pair are assumed to be depend on the identity of the base and its adjacent base pair, can be used to predict the stability of a helix (Hofacker and Stadler, 2008).

To determine the rules that govern secondary folding, small model systems that fold spontaneously can be made and their folding energies determined experimentally, using various techniques such as calorimetry and measurement of melting curves (Freier et al., 1986; He et al., 1991). In this way, thermodynamic properties of all possible base pairs with all possible neighbours have been measured, as have the stability contributions of terminal mismatches and of loops (Hickey and Turner, 1985; Freier et al., 1986; Sugimoto et al., 1987; Serra et al., 1994). These were catalogued by Serra and Turner (1995), with updated parameters published by Mathews et al. (1999), then again with chemical modifications taken into account by Mathews et al. (2004). The parameters are available for download online, along with examples on their use (Turner and Mathews, 2010). If an RNA sequence has a known sequence and structure, the free energy change of the structure compared to the unstructured RNA can be calculated by adding up the contributions of all stem, loop and terminal mismatch structures.

1.0.3 Predicting structure of a given sequence

If the calculation of the free energy change of an RNA structure compared to its unfolded form is possible when its sequence is known, then calculating the minimum free energy (MFE) structure of an arbitrary sequence becomes a computational problem. Theoretically, one could calculate the free energy of every possible structure, then pick the one with the lowest free energy, but this approach is very computationally expensive. This problem was solved by Zuker and Stiegler (1981), who showed that the global minimum free energy structure for an RNA sequence could be computed relatively efficiently using a dynamic programming algorithm.

The principle of this algorithm is to divide up the sequence into smaller ones, starting with all possible 5 nt subsequences, to calculate the energies of the subsequences, and to use these results in a dynamic algorithm to build up the larger sequence. First, the free energy of all subsequences in all possible permissible configurations (i.e., for all permissible base pairs: A-U, G-C and G-U), are computed. The first set of subsequences would therefore be nucleotides 1-5, 2-6, 3-7, and so on. The length of 5 bp is chosen because 3 bp is considered to be the smallest possible loop, so a 5 nucleotide subsequence has two possible configurations: 5 unpaired bases, or a structure with 2 paired bases and a 3 nt loop.

For a longer subsequence that goes from two arbitrary nucleotides i to j, there are two possibilities: either nucleotide j is unpaired or it is paired. If j is unpaired, the free energy for the sequence is that of the sequence i – j_{-1}, plus the energy of the 3' overhang j. As smaller subsequences are calculated before larger ones, the previously calculated value for the subsequence i – j_{-1} is used and adjusted for the 1 bp overhang.

If j is paired, it can be paired to i or to i', where the order of the nucleotides is $i < i' < j$. When paired to i, the free energy of the base i-j is added to that of the substructure $i_{+1} - j_{-1}$, which has already been calculated. When paired to i', the same process is applied to the substructure i' – j, and added to the free energy of i – i'_{-1} which has

already been computed.

This process is repeated for all subsequences one base longer, until the last one which is the whole sequence. This yields the free energy of the final sequence, but also that of all the substructures that compose it. The final structure is then built up by backtracing through the matrix of previously calculated values. This process uses experimentally measured free energy values for the small subsequences, with larger ones calculated by adding up the calculations for subsequences.

The method was adapted to permit the calculation of suboptimal structures within a certain percentage the MFE structure (Zuker, 1989), and a comparison of these suboptimal structures with the MFE structure allows an assessment of the probability that a given base pair will be present in the real structure. For example, a base pair present in the MFE and all suboptimal structures within 10% of MFE would be assigned a high score, and a base pair only present in the MFE structure would be assigned a low one.

When tested against a variety of RNA sequences up to 800 nt long, models that use this type of dynamic programming algorithm have a sensitivity of about 74%, and a positive predictive value (PPV) of 64%, that is, they identify 74% of the base pairs that form in reality, and 64% of base pairs that are identified really form (Mathews et al., 1999). Generally, the prediction tends to get worse for longer RNAs.

$$\text{Sensitivity} = \frac{\text{True positive pairs}}{\text{True positive pairs} + \text{False negative pairs}}$$

$$\text{PPV} = \frac{\text{True positive pairs}}{\text{True positive pairs} + \text{False positive pairs}}$$

The predictions of such models can be strengthened by the use of a partition function that calculates the probability that a base will be paired at equilibrium, taking into account suboptimal structures (McCaskill, 1990; Mathews, 2004). 91% of bases that pair with a probability of 99% in the predicted minimum free energy structure also exist

in the known structures derived from comparative sequence analysis. Overall, such energy minimisation algorithms can be used to predict the structure of short RNAs quite well, as well as provide an assessment of the quality of the prediction. The longer the RNA gets, the less accurate these predictions become. One important drawback of the dynamic programming algorithm presented above is that it explicitly excludes the presence of pseudoknots in the sequence (Lorenz et al., 2011). A pseudoknot is formed if any base pairs a-b and a'-b' form such that a<a'<b<b' in the sequence. Pseudoknots are known to exist in longer RNA sequences and are important in many catalytic RNAs such as ribozymes. Finally, with all its advantages and drawbacks, this method aims to calculate the minimum free energy structure, which would be expected to be the dominant structure at equilibrium. Whether this is a physiologically relevant structure will depend on the kinetics of RNA folding, as well as on interactions with other cellular components.

1.1 The sequence-structure-function relationship in RNA circuits

RNA circuits are based on a certain number of basic RNA "parts". With regard to a computational design strategy, these parts can be broadly separated into two groups: those whose function can be predicted from their secondary structure – i.e., which function by base pairing to themselves or their target –, and those whose correct function requires tertiary structure folding. The first group are particularly interesting for automatic design or RNA circuits, as their secondary structure can be predicted from their sequence. If design rules can be devised such that their function can be predicted from their structure, then we should be able to predict their function from their sequence when provided with accurate structural predictions, and therefore design new ones (Kushwaha et al., 2016). This approach has notably been used for the *de novo* design of riboregulators (Rodrigo et al., 2012; Green et al., 2014). The second group of RNA parts, that depend on specific tertiary structure, are harder

to use as tertiary structure cannot be predicted accurately based on our current knowledge. They nonetheless often contain secondary structure and some attempts can be made at incorporating within computational design pipelines, by ensuring that their predicted secondary structure is not disrupted (Shen et al., 2015), or by incorporating additional constraints into the folding calculation (Borujeni et al., 2015).

1.1.1 RNA design through abstraction of secondary structure

The correct function of riboregulators and other antisense RNAs can often be predicted based on the secondary structure calculations of their target messenger RNA (mRNA) when folded alone, or in presence of a small RNA (sRNA) regulator. This approach is based on the assumption that ribosome binding sites (RBS) are functional if unstructured, but inactive when occluded by secondary structure formation. Other constraints can also be incorporated into riboregulators, such as structural elements that enhance stability, or chaperone binding sites. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) are composed of a protein (Cas9) and guide RNAs (gRNAs), and the latter falls into this category. The gRNAs contain a region that binds to DNA through base pairing, and a stem-loop motif that specifically binds the Cas9 protein. This system is described in more detail in sections 1.2.1 and 4.1.3.

Riboregulator and antisense RNA engineering

Riboregulators are a class of RNA based regulators of gene expression. The term is a catchall used to describe small RNAs that interact with an RNA target through base pairing, whether they are found in nature or engineered. The first bacterial riboregulator to be discovered in *Escherichia coli (E. coli)* was an antisense RNA (asRNA), MicF, that directly pairs with the ribosome binding site of its target, the OmpF mRNA (Andersen et al., 1987). MicF hinders access of the ribosome binding site and inhibits translation, as well as encouraging degradation of the mRNA (Delihas and Forst, 2001). Positive antisense riboregulation was first discovered in *S. aureus* by Morfeldt et al.

(1995). Expressed alone, the *hla* gene has a self-repressing 5' untranslated region (UTR) that blocks access to the ribosome binding site. The authors showed that RNAIII could interact with the 5' of the *hla* mRNA, unfolding it and exposing the RBS. RNAIII also codes for the *hld* gene, so it is an mRNA that also acts as a riboregulator.

Both positive and negative riboregulators can be engineered. The first artificial antisense RNAs were based on the MicF RNA, redirected to pair with *lpp, ompC* and *ompA* mRNAs (Coleman et al., 1984). The synthetic riboregulators were targeted to the RBS of the mRNAs, allowing downregulation of these genes by manually changing the sRNA binding site. This approach has been shown to work in various bacterial species (Desai and Papoutsakis, 1999; Darsonval et al., 2015) and has been used to regulate natural metabolic pathways to optimise the production of chemicals of interest (Yoo et al., 2013; Na et al., 2013; Yang et al., 2015). Positive riboregulators have also been engineered (Isaacs et al., 2004; Rodrigo et al., 2012; Green et al., 2014). In the case of positive regulators, the cis-repressing 5' UTR of the mRNA also needs to be designed. These circuits were at first derived from natural riboregulator circuits (Isaacs et al., 2004). Following structural design rules from natural regulators, Rodrigo et al. (2012) created completely artificial regulators, showing that computationally encoded design rules could be used to automatically design biological parts. These first design rules focused on a 5' UTR which folded to occlude the RBS when in the off state. Green et al. (2014) devised a new set of design rules, focusing on blocking the start codon and surrounding region rather than the ribosome binding site. This allowed an improvement of the dynamic range of such circuits.

CRISPR gRNA engineering

CRISPR is a type of adaptive immune response system in prokaryotes (Barrangou et al., 2007). In type II CRISPR systems, the Cas9 protein acts as an RNA guided DNA endonuclease (Jinek et al., 2012), where 20 to 25 bases of RNA direct the nuclease to the target site on the DNA. The *S. pyogenes* CRISPR system has the advantage of using

a single Cas9 protein which uses two small RNAs – crRNA and tracrRNA – to target and cleave double stranded DNA. The crRNA contains 20 to 25 nt of complementarity to the DNA target, and requires the tracrRNA to mature and bind to the Cas9 protein. The complex then cleaves both strands of the dsDNA target. A single synthetic guide RNA can also be used, by fusing the crRNA and the tracrRNA into one short synthetic guide RNA (sgRNA, or just gRNA). The CRISPR gRNA is formed of three domains: The 20 to 25 bp of targeting region, the Cas9 handle which binds the Cas9 protein, and the *S. pyogenes* terminator at the RNA 3'. In the crystal structure, the targeting region is single stranded, with the "seed" region maintained in A-form conformation by the Cas9 protein, ready for pairing with its target DNA (Jinek et al., 2014). The Cas9 handle is formed of two stems. The first is where the crRNA and the tracrRNA are paired, and a single sgRNA can be made by linking them with a small loop. A second shorter stem loop is present further in the tracrRNA. The terminator sticks out of the Cas9 protein, with some contact with the protein in the crystal structure. The secondary structure of CRISPR gRNAs has been shown to be important for their proper function, and structured targeting regions can inhibit them (Thyme et al., 2016). CRISPR gRNAs can be incorporated within RNA circuits, by inhibiting them with antisense RNAs (Lee et al., 2016). They can also be combined with riboswitches to create small molecule responsive gRNAs (Liu et al., 2016). Additional elements such as MS2 and PP7 binding regions can also be added onto gRNAs. Zalatan et al. (2015) showed that protein effectors fused to MS2 or PP7 could be targeted to specific DNA sequences this way.

1.1.2 RNA parts which require tertiary structure interaction

RNA aptamers and riboswitches

Aptamers are short nucleic acid sequences that can specifically bind a target molecule, through interactions involving both the nucleotides and the ribose-phosphate backbone. Natural RNA aptamers that bind small molecules have been identified as part of riboswitches (Nahvi et al., 2002; Serganov and Nudler, 2013), where they serve as RNA

based small molecule sensors. Riboswitches are usually part of RNA 5' UTRs, and control mRNA translation by a conformational change upon binding, which modifies the accessibility of the RBS. Interestingly, the creation of artificial aptamers by Systematic Evolution of Ligands by Exponential Enrichement (SELEX – Ellington and Szostak, 1990; Tuerk and Gold, 1990) predates the discovery of natural ones. SELEX starts from a random library of DNA or RNA sequences, and selects for new aptamers through rounds of ligand binding in a column, followed by error-prone PCR amplification of the best binders. This technique has allowed the development of hundreds of artificial aptamers (Cruz-Toledo et al., 2012) that target small molecules, proteins or whole cells, leading to the idea that these could be used to engineer artificial riboswitches. However, the steps for transforming an aptamer into an artificial regulator have proven challenging to implement (Berens and Suess, 2015), although recent developments have shown promising results (Borujeni et al., 2015; Liu et al., 2016).

Ribozymes

Catalytically active RNAs were first discovered in the 1980s, revealing their role in complex cellular processes, beyond that of mere carriers of information (Cech et al., 1981; Guerrier-Takada et al., 1983). Many ribozymes cleave RNAs in *trans*, such as RNase P which processes transfer RNAs (tRNAs), or self-cleave, such as the Hammerhead or Hepatitis D virus (HDV) ribozymes (Serganov and Patel, 2007). Group I introns are ribozymes that autocatalytically remove themselves and ligate exons together (Cech et al., 1981; Banerjee et al., 1993). They all require tertiary interactions between the different elements of the ribozyme to be active. However, secondary structure still plays a major role in their folding, and double mutants that preserve secondary structure can often remain active even when single mutants are not (Kobori and Yokobayashi, 2016). Ribozymes can be combined with aptamers, to produce small molecule responsive ribozymes, also called aptazymes (Soukup and Breaker, 1999). They are produced by including an aptamer into one of the ribozyme loops, in such a way that the ribozyme is destabilised and rendered inactive. Upon small molecule binding, the

stem is stabilised and the ribozyme activity is restored. Similar systems that turn off upon small molecule binding can also be engineered. Aptazymes that self-cleave to release an mRNA allow small molecule dependent gene expression, by engineering a structure where ribosome access is hindered by the ribozyme, and cleavage releases the mRNA, allowing translation. Such small molecule sensors have been used in cell free extract, bacteria and eukaryotes (Ogawa and Maeda, 2007; Wieland et al., 2009; Win and Smolke, 2007). Aptazymes have also been combined with riboregulators, creating "1.5 component" signal transduction. Such systems, termed "regazymes", release a riboregulator after cleavage upon small molecule binding. The riboregulator then goes on to activate a cis-repressed mRNA, demonstrating the use of aptazymes and riboregulators in layered circuits (Shen et al., 2015).

1.2 Design rules for RNA parts

As natural RNA parts were discovered and synthetic ones created, design rules for different parts have been developed. The main rules used for creating new parts are outlined below.

1.2.1 Riboregulators, gRNAs, asRNAs, riboswitches

Positive riboregulators

Positive riboregulator design strategies are based on three main assumptions. The first is that the secondary structures of RNA strands at equilibrium can be predicted computationally, for both single strand and multistranded complexes, using programs described in section 1.0.3. The second is that initiation of translation cannot take place if the RBS and/or the start codon are obstructed by the secondary structure of the mRNA 5' region, forming the cis-repressed mRNA (cr-mRNA). The third is that the trans-activating RNA (taRNA) can interact with a cr-mRNA through solvent exposed regions, either in a linear or in a loop region. This exposed region is called

the "toehold". The design rules then aim to create a cr-mRNA / taRNA couple such that translation is obstructed when the cr-mRNA is expressed alone, but unhindered when both RNAs are expressed together. The interaction is presumed to be initiated through the toehold region, and followed by a strand displacement that leads to the new configuration. The design rules thus focus on the creation of structures which aim to maximise repression of translation in the OFF state, minimise repression in the ON state, and favour interaction between the cr-mRNA and the taRNA.

The first design rules for riboregulators were inspired by the natural *hok/sok* toxin-antitoxin system (Isaacs et al., 2004). The cr-mRNA contained a stem which occludes the RBS, and a YUNR motif (for pyrimidine-uracil-nucleotide-purine) in the loop of the stem, which creates a sharp bend in the loop and exposes its bases to the solvent (Franch et al., 1999). The cognate riboregulators initially had a similar secondary structure to the Sok RNA, with a single stranded toehold region complementary to the YUNR loop sequence. Rodrigo et al. (2012) built on this design, showing that other sRNA structures could be used, and targeting cr-mRNAs without YUNR motifs. More importantly, they abstracted the rules into secondary structure requirements that allowed unsupervised *de novo* design of taRNA/cr-mRNA pairs (see fig. 1.2).

An alternative "modular" approach that incorporated key structural changes was developed by Green et al. (2014). The authors moved away from nature-inspired structures, and created completely artificial architectures. The taRNAs, which they called "trigger RNAs", were designed to have low secondary structure, apart from a 5' stem loop for enhanced stability. The cr-mRNAs, renamed "toehold switches", contain a longer 15 nt linear toehold, at the 5' region rather than in a loop. The RBS is left unpaired in a loop, while the secondary structure formed by the stem obstructs the ATG start codon (see fig. 1.2). To allow this structure to be engineered past the start codon, the encoded protein is preceded by a linker peptide, whose sequence is determined by the bases in the second part of the stem. This means the output protein is fused to this short peptide. The toehold switches were automatically designed using

the NUPACK suite which includes an RNA design package (Zadeh et al., 2011).

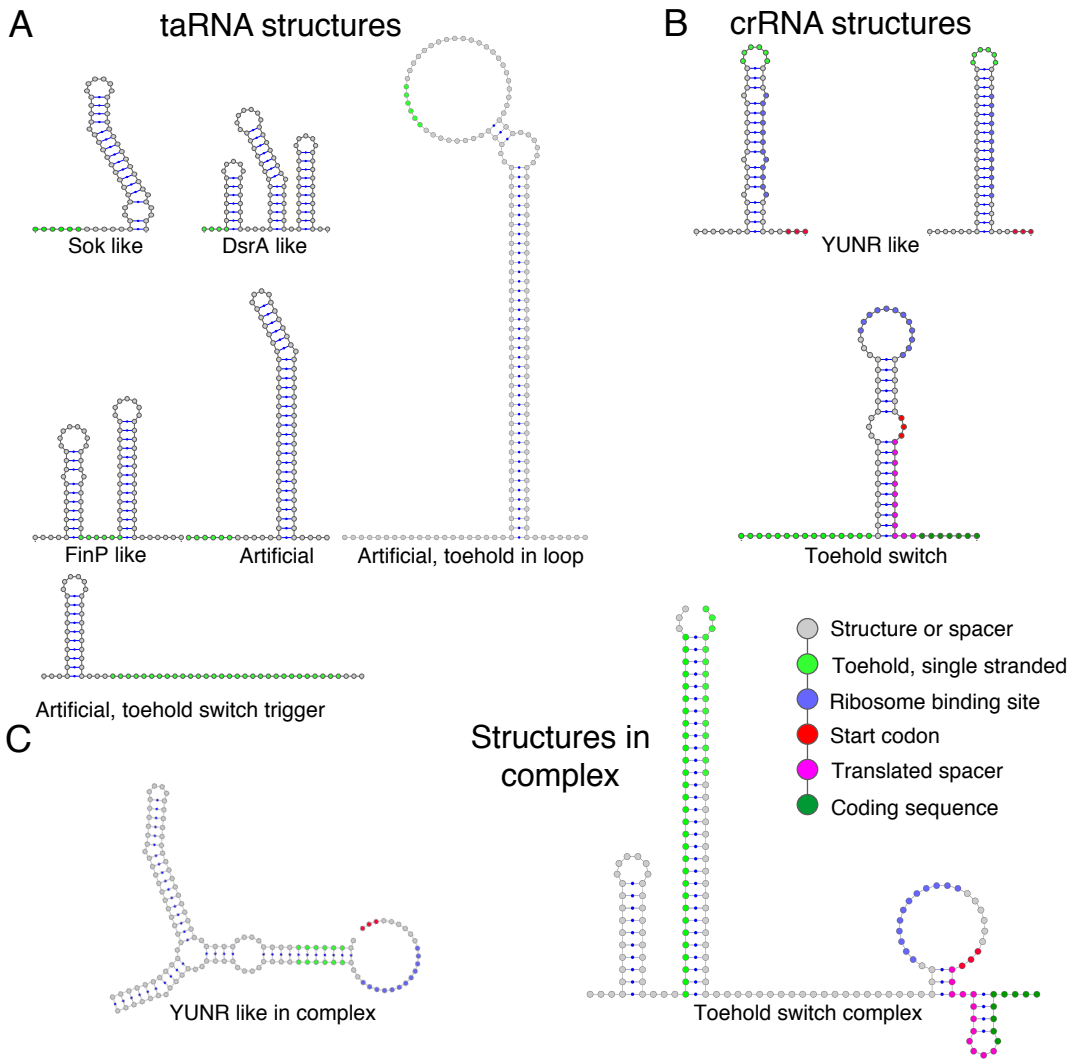


Figure 1.2 – Design rules used for riboregulator design. A. Natural and artificial riboregulator secondary structure design goals used by Rodrigo et al. (2012), and toehold switch trigger used by Green et al. (2014). B. Cis-repressed mRNA 5' UTR (with blocked ribosome binding site) and toehold switch 5' UTRs (with blocked start codon). C. Structures of taRNA and cr-mRNA complex species design goals for cr-mRNA and toehold switches.

1.2.2 Negative riboregulators

Negative riboregulators, also called simply antisense RNAs (asRNAs) were initially inspired by natural RNAs (Coleman et al., 1984). Since, different design rules have been proposed. Nakashima et al. (2006) added stem-loops at either end, creating "paired-termini asRNAs" which they found had enhanced gene repression abilities. Others added Hfq chaperone binding sites on the RNAs to improve their function (Na et al., 2013). The most thorough study was probably carried out by Lee et al. (2016), who designed and tested 96 asRNA against different targets. The authors varied target location, number of mismatches, dsRNA length, thermodynamic parameters and the presence of a YUNR motif, in *E. coli*. They found that asRNAs with dsRNA lengths over 15 nt, with a Gibbs free difference between the single species and complex species of -40 or less, and less than 15% mismatches, performed best. They also found that the presence of YUNR motifs did not result in stronger repression, and that all regions of the 5' UTR and the first 30 nucleotides of coding sequence were valid target sites. Overall their study highlighted the challenge of establishing design rules for RNAs with several important, overlapping factors, and uncovered the important factors for future asRNA design.

1.2.3 Aims of this work

This work addresses the development of design rules that allow the creation of sensors of RNA and of RNA-based actuators, using both computational and rational design. Chapter 2 presents the development of a method for creating riboregulators with a circular topology, using a Permuted Intron Exon (PIE) ribozyme. These riboregulators were first developed with computational methods. To improve their function *in vivo*, a new method for directed evolution of biological parts was tested, presented in chapter 3. Chapter 4 presents the development of sensor RNAs that use computationally designed strand-dispacement reactions that control the activity of guide RNAs from the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) system, using

the CRISPR-Cas9 system as an output. Each chapter contains relevant introductory materials, as well as a discussion and methods section. General methods are presented in chapter 5. Finally, chapter 6 discusses how the results fit into the context of RNA synthetic biology, and the potential uses of this work.

2 Circular Riboregulators

2.1 Introduction

2.1.1 Circular RNA in nature

Most studied RNAs are linear molecules, but circular RNAs (circRNAs) have been found in all domains of life (Guo et al., 2014; Danan et al., 2012; Lasda and Parker, 2014). Although they were previously considered oddities, experimental artefacts or products of spurious splicing, circRNAs have now been shown to be common (Jeck et al., 2013) and to play important regulatory roles (Memczak et al., 2013; Hansen et al., 2013; Ye et al., 2015). However, they remain relatively understudied and the reasons for their circularity remain unclear. circRNAs are found in viroids, small virus like particles composed of a single infectious non-coding circular RNA genome, and that have been suggested to be relics of pre-cellular life forms (Diener, 1989; Flores et al., 2014). Hepatitis delta virus, a satellite virus of hepatitis B virus, also has a circular genome (Macnaughton et al., 2002). In both cases, the circular RNA serves as a template for rolling circle replication, resulting in a concatamer that undergoes enzyme-mediated or autocatalytic cleavage. This process produces monomers which are then circularised to produce more virus or viroid genomes. circRNAs have also been found in some bacteria (Molina-Sánchez et al., 2006) and in many archaea

(Danan et al., 2012), where they exist as by-products of intron splicing, or in certain species as circular versions of snoRNAs and of the ribozyme RNase P (Lasda and Parker, 2014). In human cells, circular RNAs were recently shown to be more common than previously thought (Jeck et al., 2013). This has spurred much interest, especially due to the role they may play in disease (Chen, 2016a; Guarnerio et al., 2016). In humans, some circRNAs have been shown to function as miRNA 'sponges' that can suppress the activity of miRNAs, by binding to them whilst being resistant to miRNA mediated degradation (Hansen et al., 2013). However, this role as a regulator of miRNAs may be not be generalisable to other expressed circular RNAs (Guo et al., 2014), and whether many circular products have an important function or whether most are merely by-products of spurious splicing reactions remains unclear. How circularity affects RNA behaviour in general remains unexplained (Barrett and Salzman, 2016), and what roles circRNAs can play as well as the reasons for their unusual topology remains an open question.

2.1.2 Advantages of circularisation

One major change brought by circularity is the absence of free ends, which render circRNAs resistant to exonuclease degradation (Jeck et al., 2013). As some endonucleases are known to cleave close to RNA ends, they are also affected by circularisation. For example, in *E. coli*, the rapid RNA turnover – the average RNA half-life is about 5 minutes (Selinger et al., 2003; Chen et al., 2015) – is mostly due to degradation by RNase E (Mackie, 2000). Although an endonuclease, it has higher affinity for RNAs with monophosphorylated 5' ends. Dephosphorylation of newly transcribed tri-phosphorylated RNA is undertaken by the RppH protein, which tags it for subsequent degradation by RNase E (Deana et al., 2008). As a result, RNA circularisation can result in stabilisation (Mackie, 2000). circRNA transcripts have also been found to have an increased half-life compared to their corresponding linear transcripts in human cells (Enuka et al., 2015). Aside from increased resistance to nucleases, the

increased thermal stability that results from the joining of the ends has also been put forward as a possible advantage of having circular RNA. This might be important in thermophilic archaea, where several types of circular RNAs have been found (Danan et al., 2012).

2.1.3 RNA circularisation

RNA circles can be generated either through RNA back-splicing, when the acceptor splice site is situated upstream of the donor splice site (Zhang et al., 2016), or by ligation of RNA ends by an RNA ligase (Petkovic and Müller, 2015). A relatively simple way of generating artificial circular RNAs is to use self-splicing group I intron ribozymes, through the reorganisation of their sequence by a circular permutation (Puttaraju and Been, 1992). In this case the dispensable P6 loop of a T4 bacteriophage *td* gene, which contains a self-splicing group 1 intron, is removed, and the exons are joined together (see fig. 2.3). This results in a back-splicing reaction and the circularisation of the exons. Additional sequences can be inserted between the two exons of the permuted intron-exon (PIE) ribozyme, which results in their inclusion within the processed circRNA (Perriman and Ares, 1998; Umekage and Kikuchi, 2009).

2.1.4 Methods for detection of circularity

There are a number of methods used to detect circular RNA, reviewed by Jeck and Sharpless (2014, also see fig. 2.1). Divergent primer RT-PCR can detect circRNAs, as the splicing reaction reorders the primer binding sites into convergent orientations (see fig.2.1A). 2D gel electrophoresis and gel trapping are biochemical methods for detection of circular topology. The former separates the RNA in denaturing conditions, with different acrylamide concentration in each dimension. The migration speed changes linearly with acrylamide concentration, but the slope of this linear change is different for linear and circular RNAs. Linear RNAs thus appear along a diagonal on the 2D gel, whereas circular RNAs appear outside of it (fig. 2.1B). Gel trapping relies

on the mixing of RNA with melted agarose, which traps circular molecules during polymerisation. Linear molecules migrate normally, but circular molecules stay in the trap (fig. 2.1C). Exonucleases such as RNase R can also be used to degrade linear RNA, and have been used in combination with deep sequencing for high throughput circRNA detection (Jeck et al., 2013, see fig. 2.1D.).

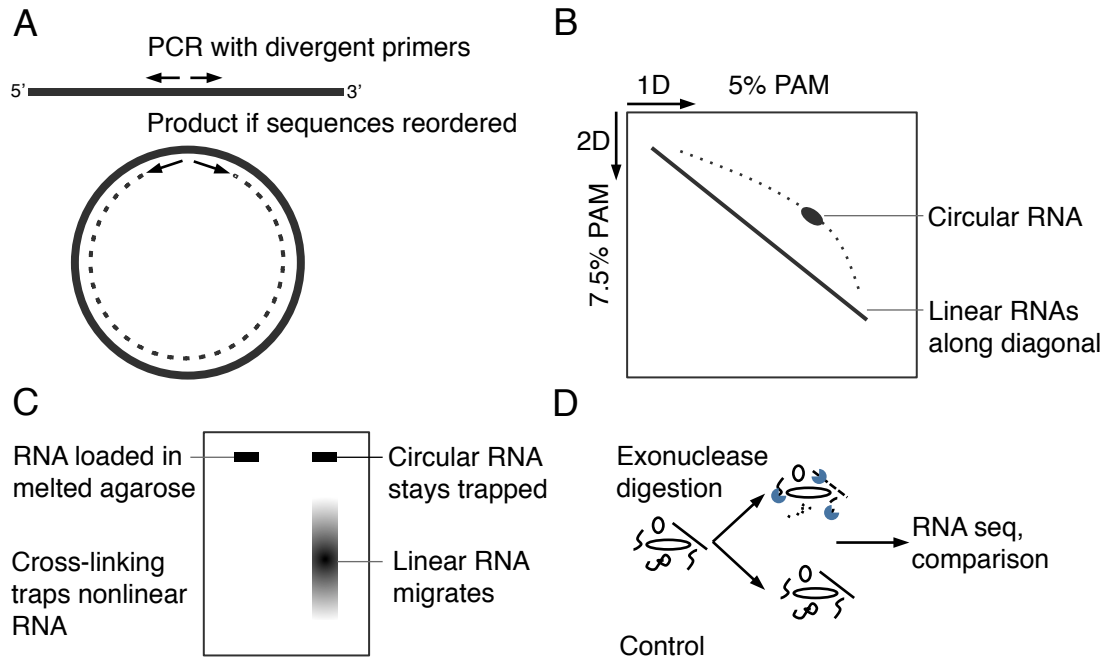


Figure 2.1 – Methods for detecting circRNA. **A**: Divergent primer PCR. Reorganisation of exons after ligation replaces primers in convergent orientation, allowing PCR amplification. **B**. 2D polyacrylamide (PAM) gel electrophoresis. Linear RNAs migrate along a diagonal. CircRNAs appear outside as they migrate more slowly in highly cross linked gels. **C**. Gel trap elecrophoresis. Non-linear RNAs are trapped by cross linking. **D**. Exonuclease such as RNase R selectively degrade linear RNA. They can be combined with next generation sequencing for high throughput circRNA detection. Figure adapted from Jeck and Sharpless (2014).

2.2 Results

2.2.1 Design of a circular riboregulator

The prevalence of circular RNAs in nature suggests that this topology may be advantageous in certain conditions. Several natural circular RNAs function as cellular regulators. Artificial circular RNAs have been produced before (Been, 1996), sometimes with a cargo included (Umekage and Kikuchi, 2009), but never as functional parts *in vivo*. Since circularisation may bring advantages such as increased thermal stability and resistance to nucleases, or simply cause different behaviour than linear RNAs within cells, it may be interesting to make use of circularised RNAs within synthetic circuits. To determine whether circular RNAs can be used as components of riboregulatory circuits, a group I self-splicing ribozyme from bacteriophage T4 with permuted introns and exons (PIE ribozyme) was used to produce circular riboregulators. The ability of the PIE ribozyme to produce circular RNAs was tested, and the effectiveness of the resulting circular RNA at activating translation from two cis-repressed mRNAs was assayed. To determine whether this is a generalisable procedure, a series of circular riboregulators were also constructed, and their ability to activate cis-repressed mRNAs compared with their linear counterparts.

To test the feasibility of using PIE ribozymes to create circular riboregulators, a previously described small RNA, taRAJ31, was used as a starting point. taRAJ31 interacts with its target through a loop-loop interaction (Rodrigo et al., 2012, see fig. 2.2), as its hybridisation region is contained in a large loop at the end of a stem. The scaffold stem of taRAJ31, which is 37 bp long, was shortened to 9 bp, yielding taRAJ31min, which was inserted within a PIE ribozyme derived from the t4 bacteriophage *td* gene self-splicing intron (Puttaraju and Been, 1992; Umekage et al., 2012).

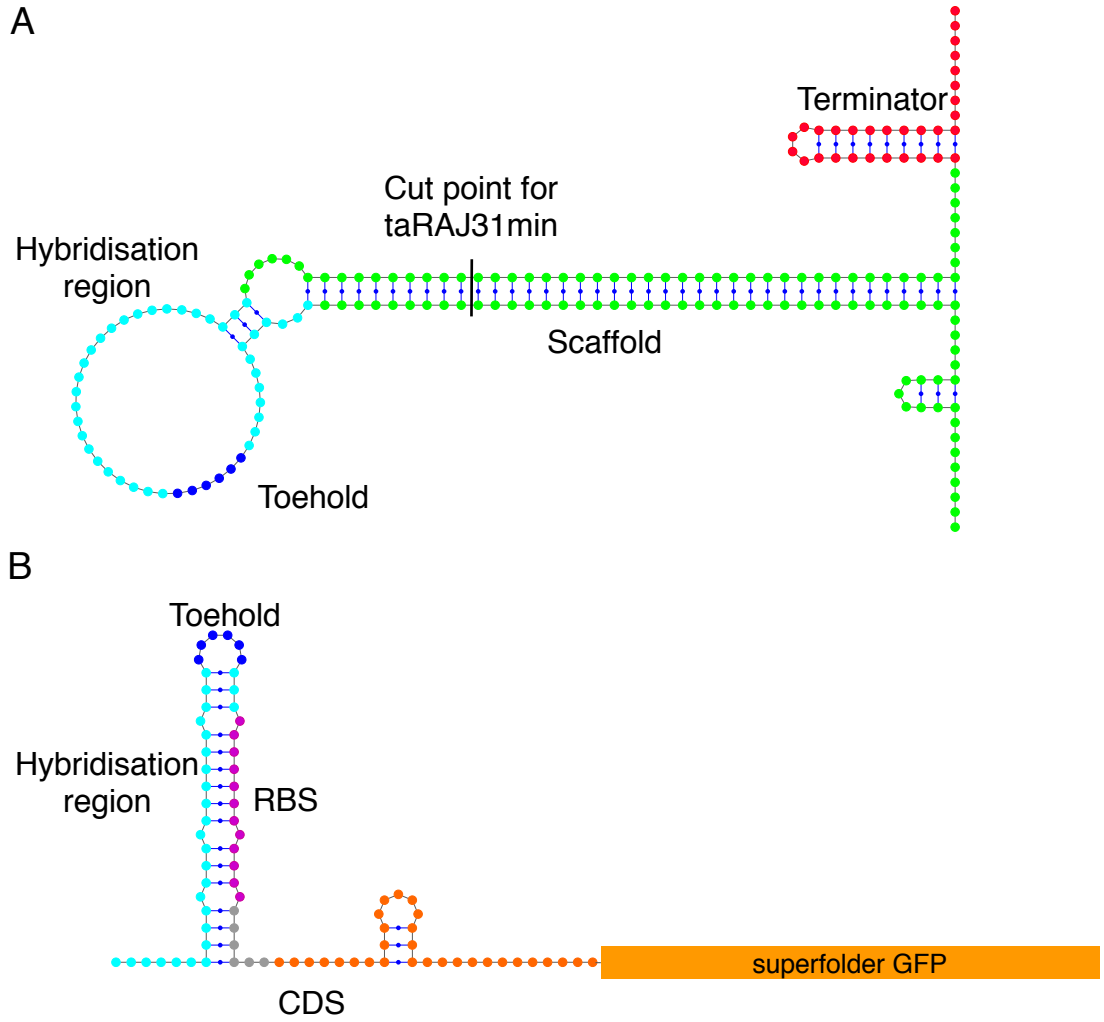


Figure 2.2 – Secondary structure of the computationally designed taRAJ31 system (Rodrigo et al., 2012), folded in NUPACK with energy parameters from Serra and Turner (1995) **A**: Structure of taRAJ31. Green bases indicate the stem that serves as scaffold. Cyan: Hybridisation region with the cis-repressed mRNA. Blue: Toehold. Red: RrnC terminator. **B**: Secondary structure of cisRAJ31 5'UTR and first 30 bases of sfGFP. Cyan: Hybridisation region. Blue: Toehold. Violet: RBS. Orange: Coding sequence.

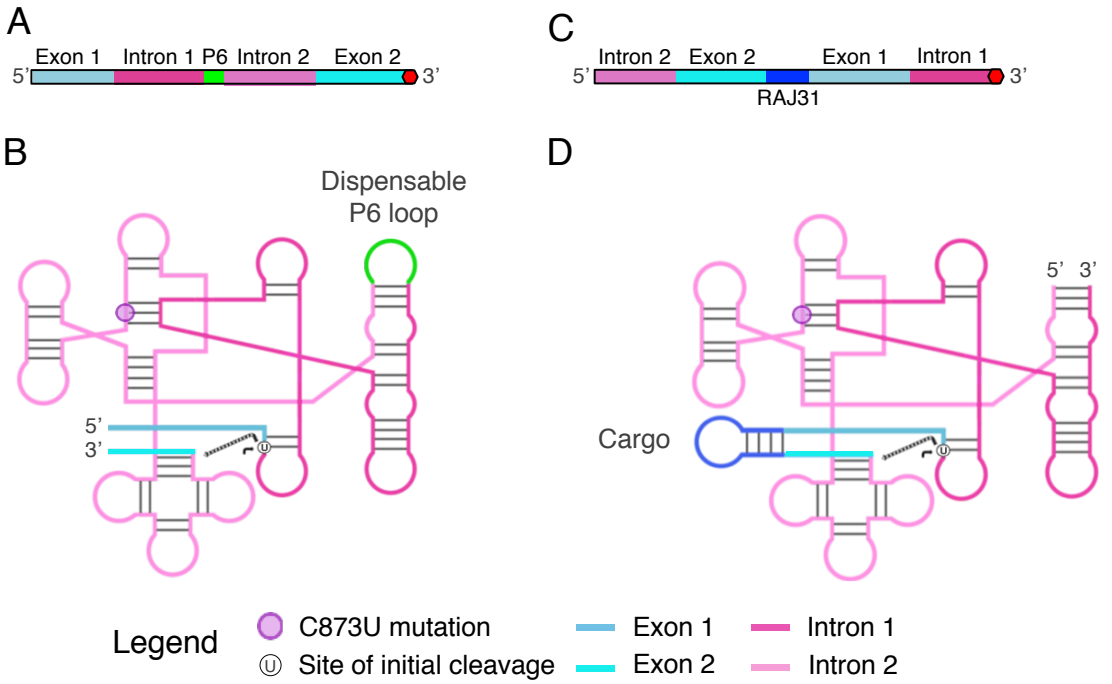


Figure 2.3 – Wild type and PIE Group 1 self-splicing ribozymes derived from the T4 *td* gene. **A**: Arrangement of the WT ribozyme at the DNA level.
B. RNA 2D structure of the WT intron, with the dispensable P6 loop in green. The splicing reaction takes place through two sequential trans-esterifications, starting with a nucleophillic attack on the indicated U. The C873U mutation greatly reduces splicing efficiency. **C**: Arrangement of PIE ribozyme at the DNA level. The gene starts with intron 2, exon 2, then exon 1 and intron 1. Cargo (here, taRAJ31min) can be inserted between the exons. **D**: Structure of the PIE intron, with the taRAJ31min riboregulator inserted as a cargo (dark blue). As the RNA splices and ligates at the same place despite the circular permutation, the taRAJ31min riboregulator becomes circularised. 2D structures were adapted from Been (1996).

The resulting sequence, circRAJ31, was inserted downstream of a $P_{L\text{-}tet}$ promoter for inducible control of expression with anhydrotetracycline (aTc). It was synthesised, and co-transformed into *E. coli* MG1655-Z1 cells, along with cisRAJ31-GFP, the cognate cis-repressed GFP (Rodrigo et al., 2012). The transcription of cisRAJ31 is controlled by a $P_{L\text{-}Lac}$ promoter, which is inducible by IPTG. MG1655-Z1 cells contain a Z1 cassette integrated onto their chromosome (Lutz and Bujard, 1997), which overexpresses the LacI and TetR repressors, thus allowing a tight control of transcription of genes expressed from $P_{L\text{-}Lac}$ and $P_{L\text{-}tet}$ using IPTG and aTc, respectively. To test whether co-expression of the circularising version of taRAJ31 could regulate translation of cisRAJ31-GFP, the cells containing both plasmids were grown with different inducer combinations, and their $OD_{\lambda=600}$ and green fluorescence was measured during exponential growth in a plate reading fluorometer. Cells grown with both IPTG and aTc, which induce cisRAJ31-GFP and circRAJ31, respectively, showed a ~2 fold increase in fluorescence than cells grown in IPTG alone (see fig. 2.4), indicating that circRAJ31 could activate translation of the cis-repressed mRNA. Although the cells showed low fluorescence in the absence of inducers or in the presence of aTc alone, cells induced with IPTG also showed an increase in fluorescence, about 7 times that of uninduced cells. This fluorescence increase in presence of cr-mRNA alone suggests leakiness of the cis-repressing 5' UTR, whereas the further increase in fluorescence when both cisRAJ31-GFP and circRAJ31 are transcribed indicated that circRAJ31 may be a functional riboregulator *in vivo*.

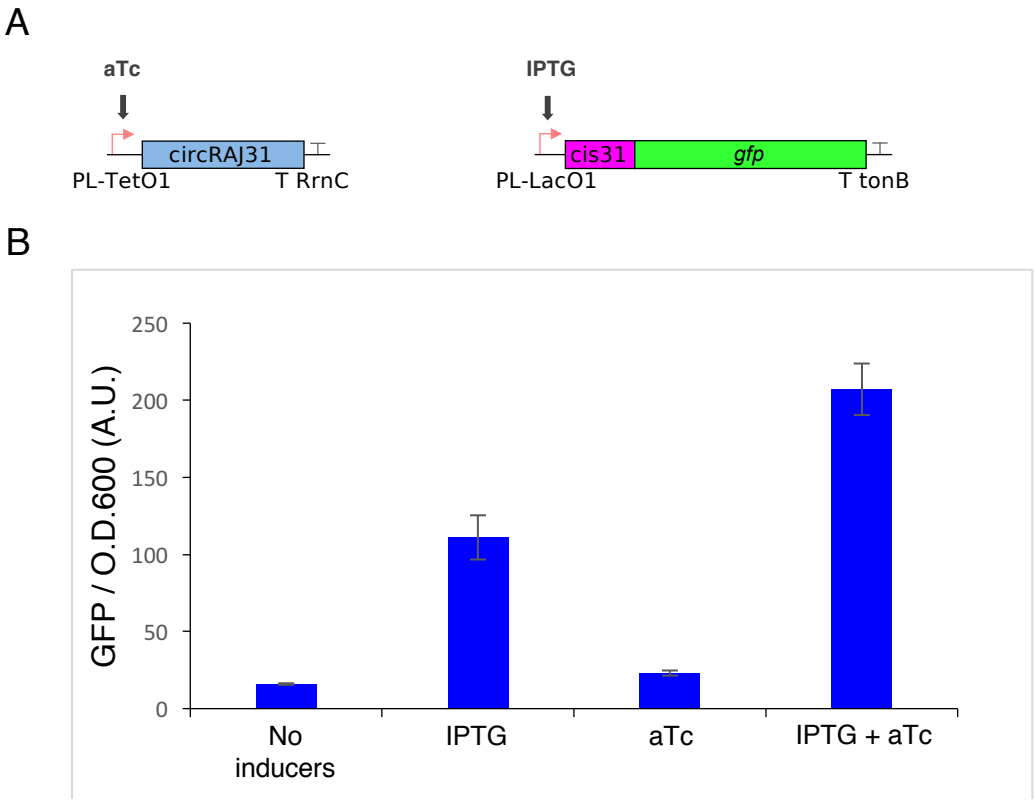


Figure 2.4 – Activation of cisRAJ31-GFP by circRAJ31. **A**. Circuit used. aTc (100 µg/mL) induces circRAJ31 transcription. IPTG (1 mM) induces cr-mRNA transcription. **B**. GFP fluorescence of *E. coli* MG1655-Z1 cells containing both cisRAJ31-GFP and circRAJ31, with different inducer combinations. Error bars represent standard deviation of three biological replicates.

2.2.2 *In vitro* RNA circularisation

RT-PCR and sequencing *in vitro* produced RNA

The ability of circRAJ31 to produce circular RNA despite the addition of the cargo was then tested. As a splicing deficient control, a C873U mutant (Belfort et al., 1987), which has very low splicing splicing efficiency (1.5% of wild type according to Brion et al. (1999)) was constructed, referred to as circRAJ31^{C873U}. Initially, both circRAJ31 and circRAJ31^{C873U} were cloned downstream of a pT7 promoter, yielding pT7-circRAJ31 pT7-circRAJ31^{C873U}, and used to produce RNA *in vitro*. After RNA extraction, an RT-PCR using divergent primers was run. Divergent primers only lead to amplification

if a splicing event changes the orientation of the target sites (Jeck and Sharpless, 2014). The products were run on a 2% agarose gel (see fig. 2.5). The PCR products from circRAJ31 separated into a fragment of the expected size (111 bp), as well as a ~222 bp band. A very faint ~333 bp band was also present on the gel. Samples from circRAJ31^{C873U} separated into a faint 111 bp band and an approximately 300 bp band. No RT controls did not lead to any amplification, a faint primer dimer band is visible on the gel. The 111 bp band confirmed the splicing of the RNA was taking place, changing the direction of the primer binding regions to a convergent orientation. The 222 bp and 333 bp bands correspond to concatamers generated during the RT step, as the reverse transcriptase repeatedly transcribes around the circular RNA template, until its RNase H domain eventually cleaves the RNA. The lanes from the circRAJ31^{C873U} also contain a 111 bp band, suggesting that some splicing takes place despite the C873U mutation, and this sample also generates circular RNA. It also contains a ~300 bp band that does not fit the size of a concatamer, and is due to a mis-splicing event (see section 2.2.2 and fig. 2.7).

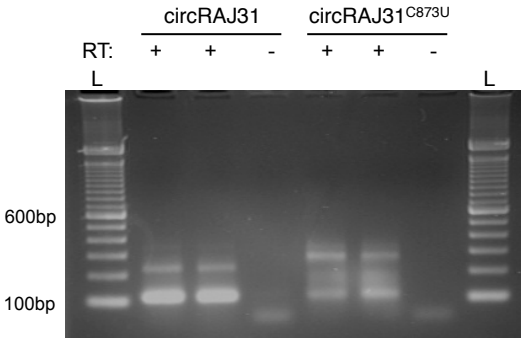


Figure 2.5 – 2% agarose gel of RT-PCR product of circRAJ31 or splicing-deficient mutant. L: Invitrogen 100 bp ladder. +: RT-PCR products of labelled samples (two repeats of RT-PCR in same conditions). -: no RT control.

Sequencing of RT-PCR products

To confirm that the splicing reaction ligated the exons at the expected site, and to investigate the nature of the unexpected bands, the obtained PCR products were

sequenced (see fig. 2.6). The sequencing of the circRAJ31 RT-PCR product confirmed the ligation site of the splicing reaction, as well as the presence of concatamers that result from the RT reaction on a circular template.

Figure 2.6 – Theoretical sequence of circularised circRAJ31, showing the ligation site between the two exons, aligned with the Sanger sequencing chromatogram of the RT-PCR products shown in Figure 2.5. The first bases are too close to the sequencing primer to be readable.

The sequencing chromatogram of the unexpected 300 bp RT-PCR product did not align with the theoretical sequence of the 111 bp circular RNA, nor with the circRAJ31 DNA. Further analysis revealed that the sequences that correspond to both exons, as well as the reverse complement of intron 1, were present. The intron 1 is located at the 5' of the exons, which normally adjacent to intro 2 in the PIE configuration. 5 out of 6 bases of the XbaI restriction site that is present just after the promoter of the PIE ribozyme are also present on the mis-spliced, just before exon td 5'. A putative 283 bp long sequence, starting with taRAJ31min, followed by the 3' exon and 3' intron, followed by the 5' exon and the rest of taRAJ31min, finishing with the circRT-fw primer, aligns well with the chromatogram trace after 90 bases (see fig. 2.7). This fragment could be generated by a mis-splicing event, where the second step of the group I intron ligation reaction attacks the wrong part of the sequence, resulting in the ligation of the exon with a part of the intron at the 5' of the RNA. Mis-splicing events are known to happen at low frequencies when the ribozyme structure is destabilised, causing cryptic splice sites to be activated and the generation of alternative splice products (Chandry and Belfort, 1987). A cryptic splice site was found to be present in intron 2, at the 5' of the

PIE ribozyme, with the spurious splice product only present with the C873U mutation. This sequencing result suggests that the C873U mutation destabilised the ribozyme, but without disrupting it entirely – with some splicing still detectable – and whilst also causing some mis-splicing reactions due to the destabilised ribozyme structure.

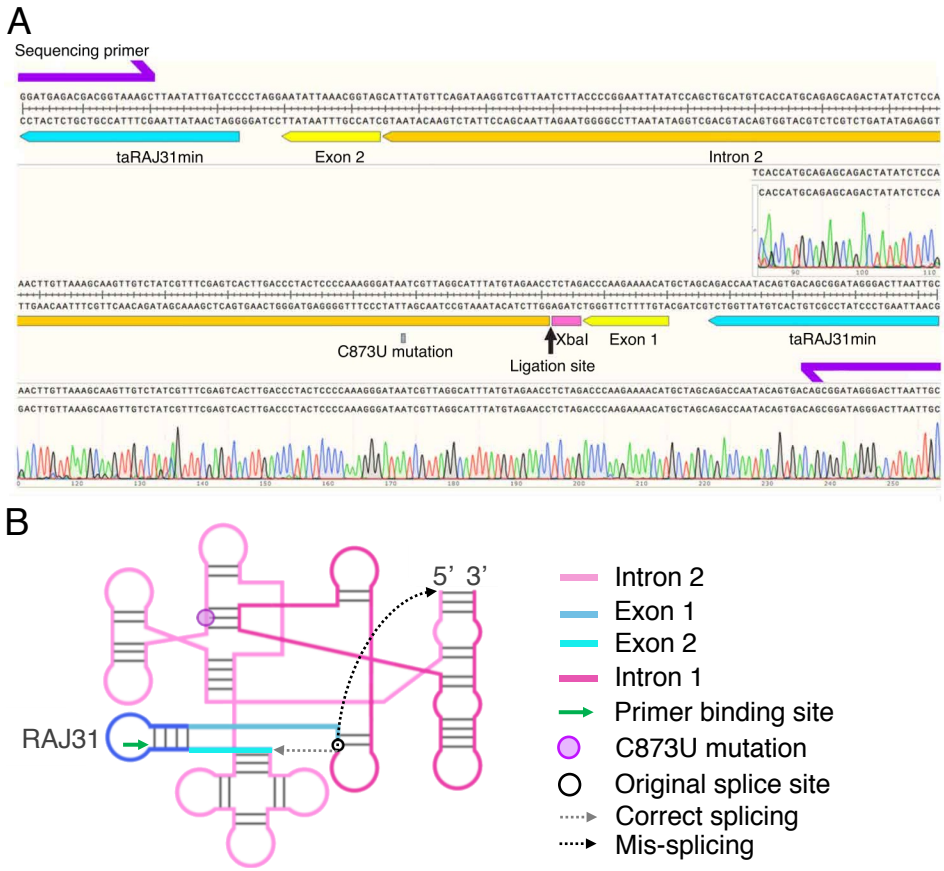


Figure 2.7 – **A**: Alignment of the sequencing chromatogram of the RT-PCR with the putative sequence that would result from the mis-splicing. **B**: Mis-splicing event where intron 2 attacks the splice site.

2.2.3 *In vivo* RNA circularisation

With the confirmation that the introduction of the taRAJ31min riboregulator between the two exons of a PIE ribozyme led to the generation of a circular RNA containing the riboregulator, and the establishment of a method to detect the circular RNA, total RNA was extracted from *E. coli* expressing either circRAJ31 or circRAJ31^{C873U}. The RT-PCR

was repeated with both convergent PCR primers (to detect the unspliced RNA) and divergent primers (to detect the circular RNA, see fig. 2.8). A band of ~111 bp is seen in both cases with primers designed to amplify the linear RNA, but only in samples from circRAJ31 when divergent primers were used. The approximately 300 bp band from the mis-splicing event detected from *in vitro* transcribed samples was not visible. PCR products were sequenced, confirming the reorientation of exons and the correct splice site in circRAJ31 samples, as in fig. 2.6. Amplicons from the linear samples showed the presence of the introns. The PCR product from circRAJ31^{C873U} after amplification with divergent primers was sequenced, even though no product was visible on the gel, and did not yield a readable chromatogram. Neither the circular RNA or the mis-spliced RNA that were detected after *in vitro* expression were detectable when the procedure was carried out with RNA produced in *E. coli*.

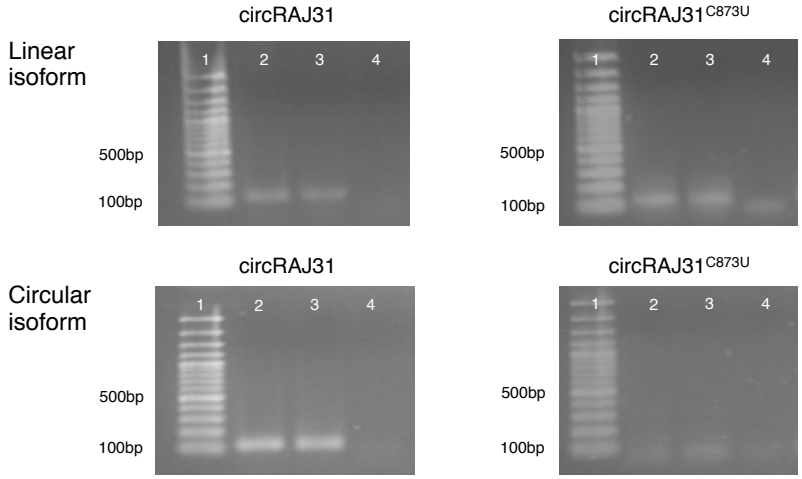


Figure 2.8 – Agarose gels of RT-PCR of *E. coli* total RNA, expressing either circRAJ31 or circRAJ31^{C873U}. RNA was extracted from *E. coli* expressing either circRAJ31 (left column) or circRAJ31^{C873U} (right column), and the PCR was carried out using either convergent primers specific to the unspliced product (top row) or divergent primers specific to the circular product (bottom row).

2.2.4 Splicing is required for circRAJ31 activation of translation

To reduce the background of GFP that is visible with untagged GFP, an LAA degradation tag (Keiler et al., 1996) was added to its C terminus. The LAA peptide tags the protein

for degradation by the ClpX protease machinery, significantly shortening its half-life (Andersen et al., 1998). The ability of circRAJ31 and circRAJ31^{C873U} to activate translation of cisRAJ31-GFP-LAA was then compared. As the unspliced RNA also contains the taRAJ31min sequence, unspliced RNA products may also activate cis-repressed mRNAs if the hybridisation sequence is solvent exposed and able to bind the cr-mRNA target. DH5α cells containing cisRAJ31-GFP-LAA and either circRAJ31 or circRAJ31^{C873U} were grown in a fluorometer in the presence or absence of IPTG and aTc for induction of cis-repressed mRNA and of the riboregulator, respectively, and their fluorescence during exponential phase measured (see fig. 2.9A). In the absence of inducers, or in the presence of a single inducer, fluorescence was low for both constructs. In the presence of both aTc and IPTG, the fluorescence of cells containing the circular riboregulator and cr-mRNA was about 22 times that of cells induced with IPTG alone, whereas expression of circRAJ31^{C873U} only increased fluorescence by about 2 fold. This suggests that the splicing is required for efficient activation of cisRAJ13, and therefore that translational regulation by a circular riboregulator is possible (see fig. 2.9B). To confirm that circRAJ31 was activating translation through a toehold-initiated base pairing mechanism, a mutant with 4 mutations in the toehold binding region of circRAJ31 was constructed, called circRAJ31^{Toe}. These mutations completely abolished GFP activation (see fig. 2.9C), confirming the importance of the toehold region.

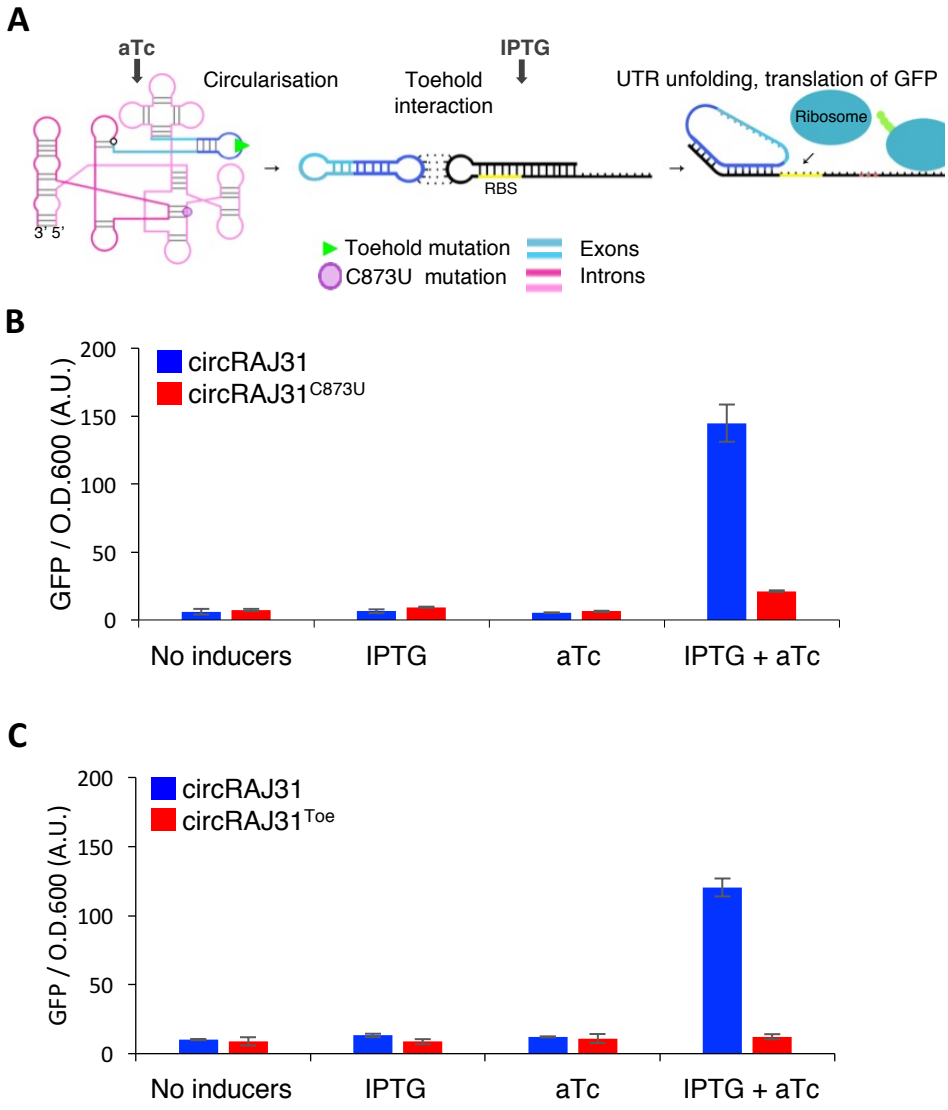


Figure 2.9 – Activation of cis-repressed GFP by circRAJ31. **A**. Circuit used. **B**: Fluorescence of *E. coli* DH5α containing cisRAJ31-GFP-LAA and either circRAJ31 or circRAJ31^{C873U}, with different inducer combinations indicated on the x axis. **C**. Fluorescence of cells containing cisRAJ31-GFP-LAA and either circRAJ31 or circRAJ31^{Toe}, with different inducer combinations indicated on the x axis. circRAJ31, circRAJ31^{C873U}, circRAJ31^{Toe} are under the control of an aTc inducible promoter, cisRAJ31-GFP-LAA is under the control of an IPTG inducible promoter. Error bars represent the standard deviation of three biological replicates. IPTG concentration: 250 µM, aTc concentration: 100 ng/mL.

2.2.5 Circular riboregulator can control antibiotic resistance gene expression

To test the possibility of using an antibiotic resistance marker to select improved versions of the circular riboregulator, or to create new ones, the *cat* gene was cloned downstream of the cisRAJ31 5' UTR, generating cisRAJ31-*cat*. cisRAJ31-*cat* and either circRAJ31 or circRAJ31^{C873U} were co-transformed in *E. coli* DH5α cells, then plated on agar plates prepared with different combinations of inducers and antibiotics. Control plates did not contain any chloramphenicol, whereas others contained chloramphenicol and either no inducers, aTc, IPTG, or aTc and IPTG. At IPTG concentrations of 250 µM and/or aTc concentrations of 100 ng/mL, cells containing circRAJ31 and cisRAJ31-*cat* only grew on plates containing both IPTG and aTc (see fig. 2.10). At higher concentrations of IPTG or with longer growth times, colonies also grew on IPTG-containing plates, which suggests leakiness of the cis-repressing UTR when controlling the *cat* gene.

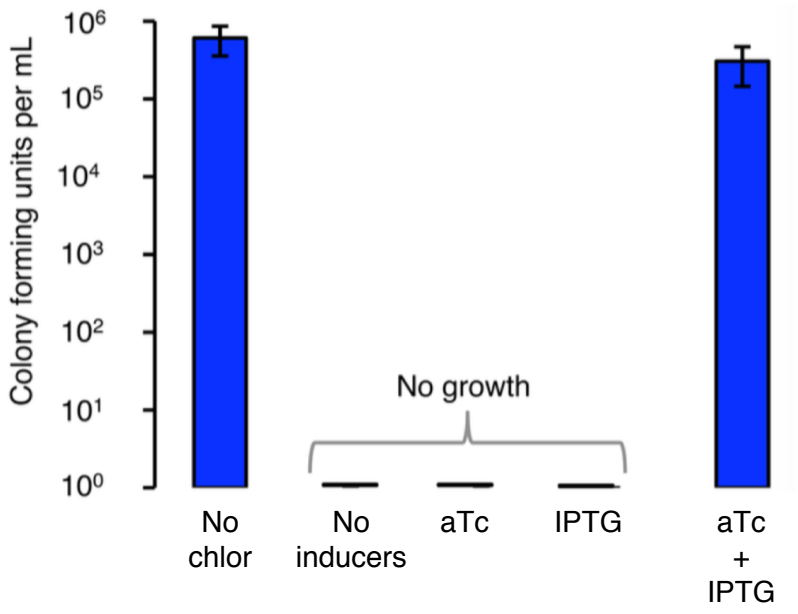


Figure 2.10 – CFU per mL of *E. coli* DH5α cells transformed with circRAJ31 and cisRAJ31-*cat* on agar plates containing chlorampenicol (chlor) and different inducers. IPTG and aTc control induction of cisRAJ31-*cat* and circRAJ31, respectively. IPTG: 250 µM, aTc: 100 ng/mL.

The cis-repressed *cat* gene can also be controlled by circRAJ31 in liquid culture. *E. coli* DH5α cells containing this circuit grow faster in LB containing chloramphenicol if IPTG and aTc are also present (see fig. 2.11). However, this is only the case at certain concentrations of chloramphenicol. In LB containing 35 µg/mL of chloramphenicol, samples containing IPTG grow at the same speed as samples containing both IPTG and aTc. At high concentrations (70 µg/mL), samples where circRAJ31 is induced by aTc grow faster than uninduced samples (see fig. 2.11).

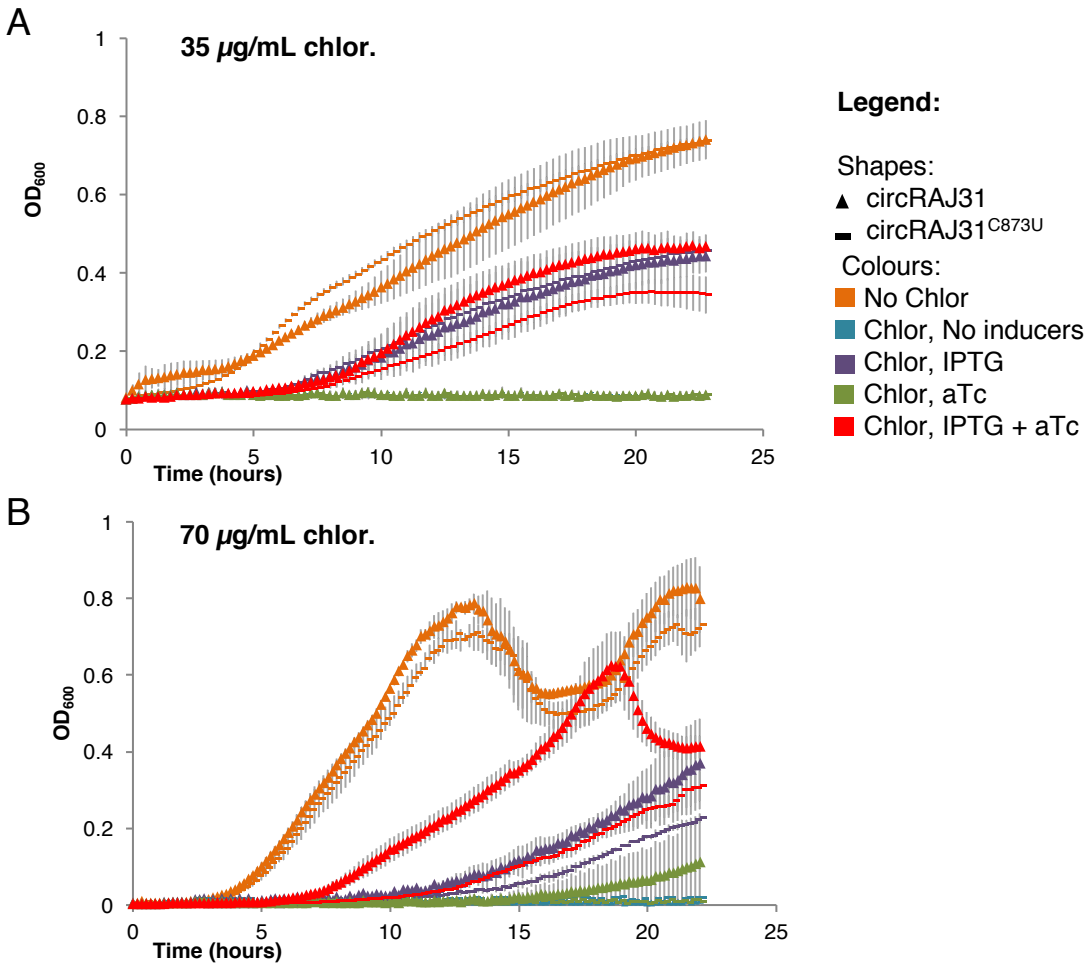


Figure 2.11 – Liquid cultures of *E. coli* DH5α in M9 containing different concentrations of chloramphenicol and inducers. **A**: 35 µg/mL chloramphenicol. **B**: 70 µg/mL chloramphenicol. Triangles show circRAJ31 samples, rectangles circRAJ31^{C873U} samples. Colour codes show inducer combinations (IPTG, 250 µM, and aTc, 100 ng/µL). Error bars represent standard deviation of three biological replicates.

2.2.6 Comparison of circRAJ31 with a linear version

To investigate how the circular riboregulator compared to a linear one for trans-activation, the linear equivalent of circRAJ31 (taRAJ31min) was synthesised and cloned into the same plasmid as circRAJ31. taRAJ31min was found to be a better activator of cisRAJ31-GFP-LAA, with a 65% higher level of fluorescence over $OD_{\lambda=600}$ (see fig.2.12).

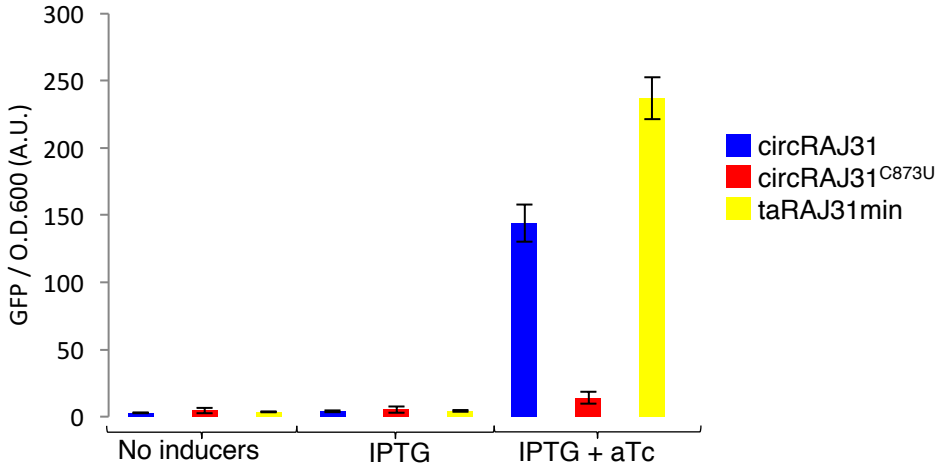


Figure 2.12 – Fluorescence of *E. coli* DH5α cells with cisRAJ31 under the control of an IPTG inducible promoter, and either circRAJ31, circRAJ31^{C873U} or taRAJ31min, under the control of an aTc inducible promoter. IPTG: 250 µM, aTc: 100 ng/ µL . Error bars represent standard deviation of three biological replicates.

2.2.7 Engineering new cis-repressed mRNAs

Although circRAJ31 and taRAJ31min could activate translation of the cisRAJ31 mRNA, the relative difference between samples expressing mRNA alone and those co-expressing the mRNA and circRNA activator remained lower than some published RNA regulators (Green et al., 2014). Growth of samples expressing cis-repressed *cat* without any RNA activator suggested that the cis31 cr-mRNA structure was leaky. To improve this, new cis-repressed RNAs were constructed according to new design rules (Green et al., 2014). New cis-repressed messenger RNAs that can be activated by circRAJ31 according to new secondary structure rules were first designed using a custom NU-

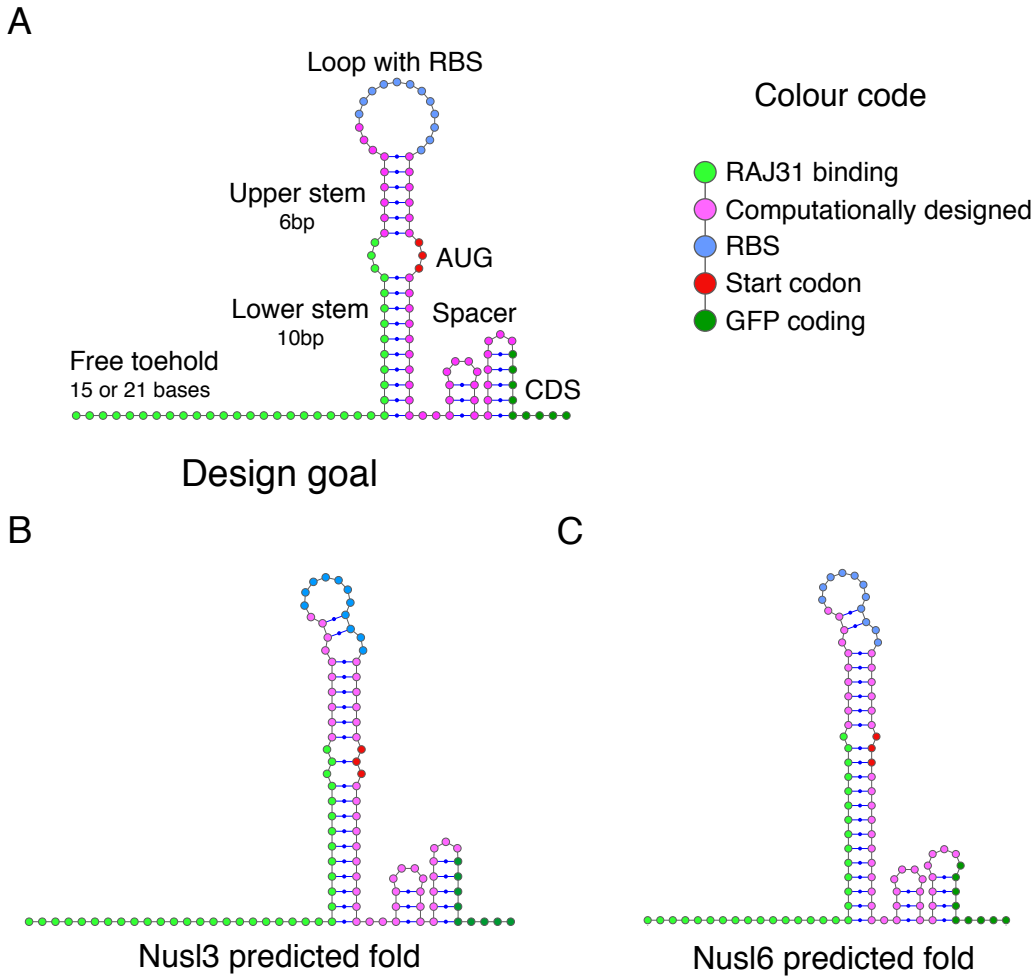


Figure 2.13 – Design of reverse engineered cis-repressed mRNAs using a design goal based on (Green et al., 2014) and evaluated using the sRNA calculator (Borujeni and Salis, 2011).

PACK script (Dirks et al., 2004; Zadeh et al., 2011; Wolfe and Pierce, 2014). The new design goal included a long single stranded toehold at the 5', an exposed RBS in a loop, but with strong secondary structure around the start codon (see fig. 2.13 for exact structural constraints). As the RAJ31min hybridisation sequence is longer than that of the trigger RNAs (31 vs 25), two alternative designs were made: One with an extended 21 nt toehold, and one with a 15 nt toehold, the same as normal toehold switches. The 10 best design sequences were curated to remove those that contained stop codons in the coding sequence, then were evaluated using the alpha version of the Salis lab sRNA calculator available on the Salis lab website (Borujeni and Salis, 2011). The sRNA

calculator is a variant of the RBS calculator (Salis et al., 2009), and predicts changes in translation initiation rate of an mRNA 5' UTR after sRNA binding. The highest scoring sequence for each design were named Nusl_3 (with a 21 nt toehold) and Nusl_6 (15 nt toehold). They were synthesised, co-transformed into *E. coli* cells and characterised in the presence of circRAJ31, circRAJ31^{C873U} and the linear taRAJ31min. The newly RAJ31 activated cr-mRNAs showed no significant activation in the presence of circRAJ31. However, the increase in fluorescence in the presence of both Nusl_3 and Nusl_6 by taRAJ31min was high (approximately 10 fold and 32 fold, for Nusl_3 and Nusl_6, respectively – see fig. 2.14).

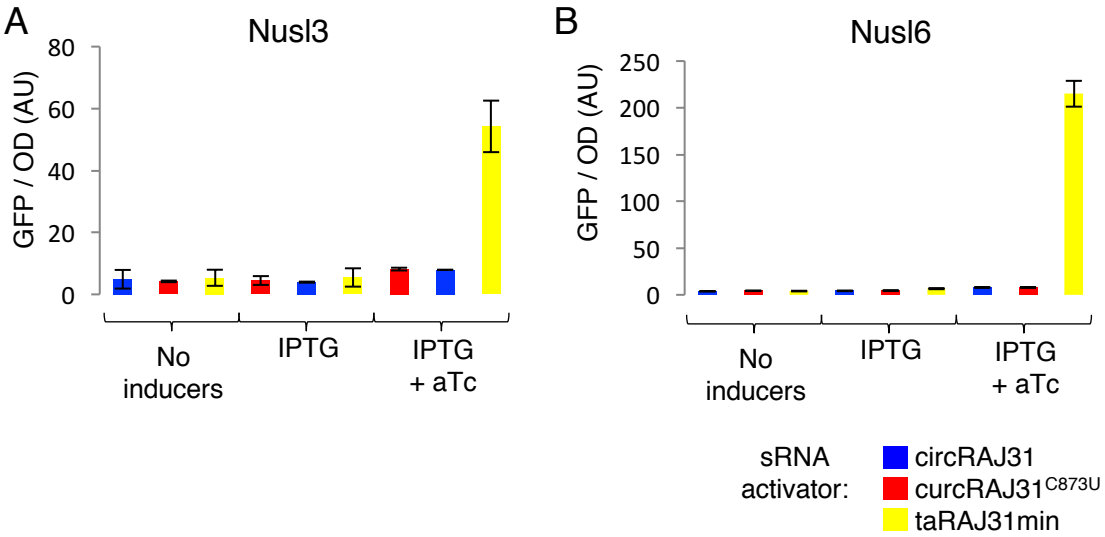


Figure 2.14 – Fluorescence of *E. coli* DH5α cells expressing a circRAJ31, or taRAJ31min, and 2 variants of cis-repressing 5' UTRs reverse-engineered to be activated by RAJ31, according to the Green et al. (2014) design rules. **A**: Nusl3 and **B**: Nusl6. Error bars represent standard deviation of three biological replicates.

2.2.8 New circular riboregulators

The circRAJ31 riboregulator was found to be effectively circularised by the PIE ribozyme, although it had a marginally lower activation fold than taRAJ31min. To find out whether PIE circularisation is a generalisable strategy for circularising riboregulators, and to further investigate the differences of circular riboregulators with their linear counterparts, 5 more variants of circular riboregulators were designed

and synthesised, and tested against their corresponding cis-repressed mRNAs. The new circular riboregulators were engineered to activate previously described 'toe-hold switch' cr-mRNAs (Green et al., 2014). They all contain the same circularising ribozyme sequence as circRAJ31, but each has a different hybridisation region inserted between the exons (the hybridisation region corresponds to the "cargo" in figure 2.3D, sequences are given in appendix B). For each regulator, a C873U splicing mutant and a linear riboregulator was generated to compare their activity (see fig. 2.15). Riboregulators circRAG1v1, circRAG1v2, circRAG2v2 and circRAG4v1 were found to increase fluorescence of their cognate cis-repressed mRNAs.

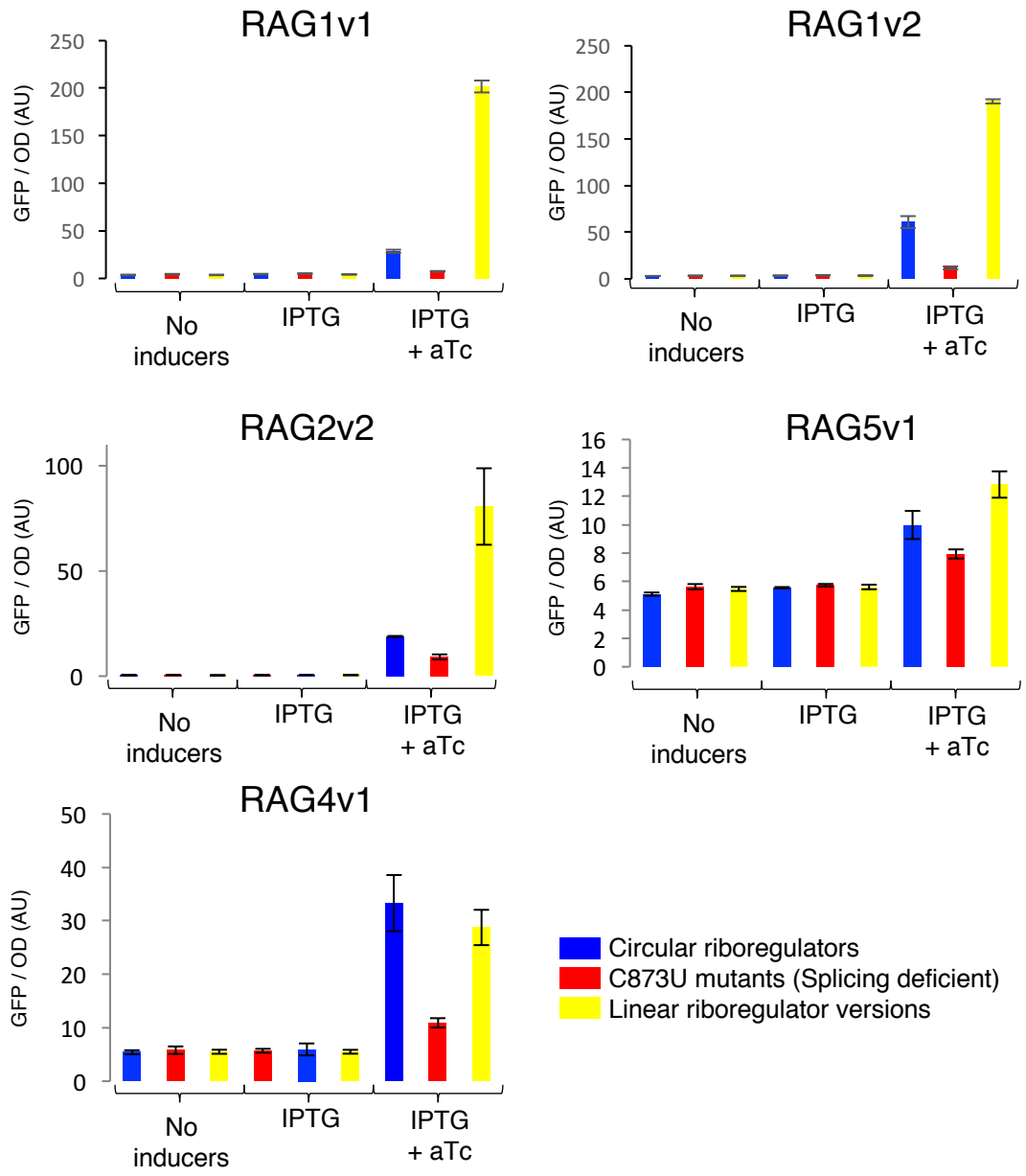


Figure 2.15 – Fluorescence of *E. coli* DH5 α cells expressing cis-repressing 5' UTRs, and their cognate riboregulators. UTRs used are indicated in the title of each graph. Blue: circRAJ31, Red: Non-splicing C873U mutants. Yellow: linear versions from (Green et al., 2014). Error bars represent standard deviation of three biological replicates.

However, the circular regulators never activated translation better than their linear counterparts. The fluorescence of cells expressing the riboregulators circRAG1v1, circRAG1v2 and circRAG2v2 was much lower than that of cells expressing their linear counterparts. Like circRAJ31 and taRAJ31min, circRAG4v1 and RAG4 have similar activation folds. In the case of RAG5v1 regulators, the circular regulator was not found to activate translation and the linear regulator had a marginal activation fold of ~2 fold. The calculated activation folds for all parts are shown in figure 2.16.

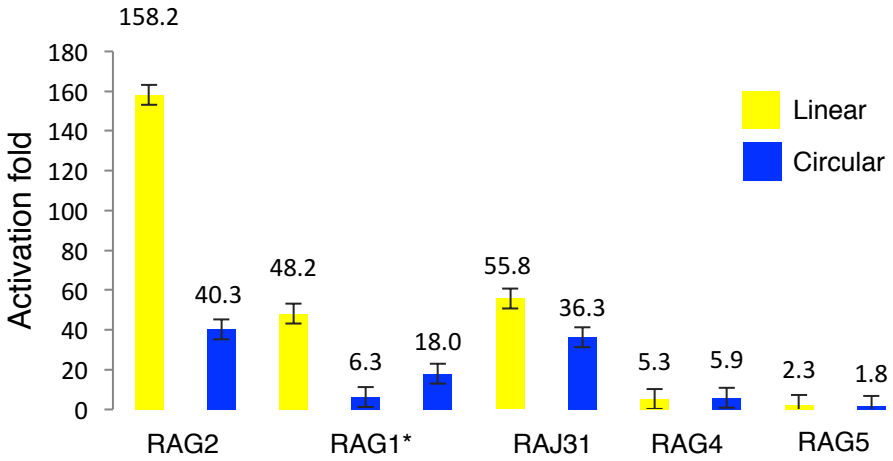


Figure 2.16 – Apparent activation folds of cis-repressed GFP mRNAs (indicated on the x axis) by linear versions (yellow) or circular versions (blue) of riboregulators. Two different versions of circRAG1 were tested. Values were calculated by dividing fluorescence values when both activator and cis-repressed mRNA are induced, compared those when the cr-mRNA is alone. Error bars represent standard deviation of three biological replicates.

2.3 Discussion

2.3.1 PIE ribozymes for synthetic regulatory circRNA production

The incorporation of riboregulators within PIE group I ribozymes enabled the generation of circular sRNA regulators. The ribozymes splice correctly despite the addition of a sRNA cargo between the exon sequences, and can be used to produce circular sRNAs *in vivo* as well as *in vitro*. In 4 out of the 5 cases tested, the introduction of the cargo did not impede circularisation, which was either observed directly, or indirectly through the introduction of the C873U mutation, which disrupts ribozyme functionality. This mutation reduced or abolished riboregulator activity. RNA-RNA interactions, including those involving natural circular RNAs, are known to interact through base pairing. However, for riboregulators to be effective, they need to be present in sufficient quantities to activate their cognate 5' UTR targets. The low splicing efficiency of some ribozymes had been noted (Olson and Müller, 2012), but our results show that PIE group I intron ribozymes can produce enough circRNAs to be used as *in vivo* regulators of gene expression. Circular riboregulators were able to regulate translation from various cis-repressed 5' UTRs, for two different cis-repressed mRNAs coding for a fluorescent protein and for an antibiotic resistance gene. Mutations in the 'toehold' loop region of circRAJ31 abolished riboregulator activity, suggesting that like natural riboregulators, our circular riboregulator initially interacts with its target through base pairing of solvent exposed regions. Circular riboregulators, like linear ones, then modify the 2D structure of their target cis-repressed mRNA 5' UTRs *in vivo*, and can be incorporated as actuator parts in synthetic RNA circuits. PIE ribozyme circularisation is easy to achieve and is an effective method to produce circular riboregulators *in vivo*. To our knowledge, this is the first example of a synthetic circular sRNA regulator being used to regulate gene expression in living cells.

2.3.2 Advantages and drawbacks of PIE circularisation

Generation of PIE-ribozyme circularised RNAs is extremely straightforward as it simply requires inclusion of the desired sequence between the two exons of the sequence. Out of 6 PIE ribozyme encased riboregulators tested, 5 were more active than the non-splicing mutant used as a control, supporting the case that PIE effectively circularises these riboregulators and allows them to interact with their targets. The 6th (circRAG5v1) was not more active than its non-splicing control. However, its linear equivalent linRAG5 only had a marginal activation fold of 2.3, compared to ~450 when tested by Green et al. (2014), albeit in a different genetic context. This suggests that cisRAG5 is not a good target for riboregulators in this genetic context, and does not give us any information on whether circRAG5 is circularised or not. The PIE ribozyme method has also been used by others for producing various kinds of circular RNAs. Perriman and Ares (1998) used it to create circularised mRNA *in vivo*, and Umekage and Kikuchi (2009) to create large amounts of circular aptamers for purification. The production of synthetic circular riboregulators extends this list, and suggests that PIE ribozyme circularisation is a robust method for producing many different synthetic circular RNA *in vivo*, despite the addition of various 'cargo' between the two exons.

It is possible however that certain sequences would disrupt ribozyme activity and thus RNA circularisation. For example, a cargo which has strong homology to the ribozyme sequence may disrupt its secondary structure, and thus the formation of the catalytic core, blocking circRNA production. In these cases, a simple potential solution would be to circularise said sequences using a PIE ribozyme from another organism, such as *Tetrahymena*. Many group I self-splicing introns have been identified, and they show little conservation at the sequence level (Nielsen and Johansen, 2009), even though they have conserved secondary structure. It should therefore be possible to find a group I intron without homology to the cargo and circularise such sequences in this way. Other methods for circularising RNA, such as via an RNA ligase (Englert and Beier, 2005; Petkovic and Müller, 2015), could also be used, although it is unknown

whether they will produce sufficient quantities of circRNA for them to be used as regulators.

2.3.3 Activation of translation by unspliced RNA

It is noteworthy that several C873U mutants also activated translation of cis-repressed mRNA. Circularisation is therefore not required for activity of these parts, which is to be expected as previously published riboregulators are linear. Rather, low activation by non-splicing mutants may be due to inaccessibility of the hybridisation region of the unspliced PIE-riboregulator RNA, as can be the case when cis-repressed UTRs interact with mRNAs (Green et al., 2014). The ribozyme is a large molecule (383 bp) compared to the the spliced riboregulator (111 bp), and has strong secondary and tertiary structure, which could interfere with RNA-RNA interactions unless removed. The fact that circRAG1v2 had a higher activation fold than circRAG1v1, for both the circularising version and the C873U mutant, tends to support this hypothesis. The hybridisation regions of both riboregulators are identical, but circRAG1v2 has a longer stem region than circRAG1v1. This may allow the taRAG1 cargo to extend out of the ribozyme structure and activate its target even when uncleaved. This is also supported by the strong activation of both Nusl_3/Nusl_6, and of 'RAG' toehold switches, by linear RNAs. The mis-splicing event that is detected in the *in vitro* produced samples (but not *in vivo*) also adds some complexity to using ribozymes for circular sRNA production. Mis-splicing events are known to occur in destabilised ribozymes (Chandry and Belfort, 1987), and in this case, a cryptic splice site seems to have been reactivated by the C873U mutation. The mis-spliced RNA was not detected *in vivo*, and mutant had lower activation levels than the functional ribozymes. However, this mis-splicing event also raises the possibility of other spurious splicing products being generated at very low levels. Although these rare products would not be an issue for our assay, they may be for other applications of circular RNA. Interestingly, it is also through mis-splicing that many natural circRNAs are thought

to be created, although how many of these are accidental, spurious products and how many actually play a useful role remains a point of debate (Guo et al., 2014). Another possibility for higher activity of circRAJ31 versus circRAJ31^{C873U} is that the circular products accumulate to higher levels in the cells than the linear, splicing deficient mutant. Circular RNA is known to have high stability in *E. coli* cells (Mackie, 2000). Even if this is not an advantage in this context of fast growing *E. coli* cells, which are characterised during exponential growth phase, it may be interesting in other cases, discussed further on.

2.3.4 Differences between linear and circular RNA

In all cases, the circular versions of the riboregulators had similar or lower activation folds than their linear counterparts, and inclusion within a PIE ribozyme for circularisation did not increase the activation fold of the riboregulators tested. This could be due to a relatively low splicing efficiency of the riboregulators in this context, leading to most RNA remaining unspliced, or to the circular RNA being less effective in regulating cis-repressed mRNA. Although the goal of this project was to determine whether circular RNAs could be used in bacteria, testing purified circular riboregulators *in vitro*, for example in a transcription-translation cell free system (Sun et al., 2013), would help our understanding of the differences when using circular rather than linear RNA parts. Lower activation folds *in vitro* at identical activating riboregulator concentrations would suggest that the constraints brought by circularisation impede interaction with cr-mRNAs. Similar or higher activation by circRNAs would suggest that the lower *in vivo* levels seen might be due to low splicing efficiency or differences in stability, which were not determined in this work.

2.3.5 Design rules for synthetic circRNAs

Regardless of the lower activation fold by circular RNAs, it is clear from these results that the design rules used to produce linear taRNA/cr-mRNA couples cannot be

used directly with circular regulators. The Nusl3 and Nusl6 cr-mRNAs (fig. 2.14), that were reverse-engineered to be activated by taRAJ31min according to the design rules of Green et al. (2014), showed a large increase in fluorescence in the presence of taRAJ31min, but none by circRAJ31, even though both RNAs contain the same hybridisation sequence. The originally tested cisRAJ31 cr-mRNA, designed by Rodrigo et al. (2012), has the same hybridisation region but a different 2D structure, and is activated by both the linear and circular versions of RAJ31. The circRAG4 regulator also had a similar activation fold as linRAG4, whereas linRAG2 and linRAG1 had much higher activation folds than circRAG2 and circRAG1, respectively. Although these experiments are too few to determine what structure of cr-mRNA would interact better with circRNAs, new rules could be determined if a larger number of cr-mRNA structures were tested. Nonetheless, some the rules for riboregulator design remain valid. The mutation in the toehold totally abolished circRAJ31 activation of cisRAJ31, despite 34 bases of complete homology between the taRNA and the cr-mRNA target, showing that the solvent exposed toehold loop of the cr-mRNA is an important region for the interaction between RNAs and comforting the hypothesis that hybridisation is established through solvent exposed regions of the RNAs. In the case of RAJ31, this is through a kissing loop interaction, whereas for the RAG series of regulators, the toehold of the cr-mRNA is in the solvent exposed 5' linear region.

2.3.6 circRNA containing circuits and RNA structure prediction

All of the linear versions of these riboregulators were computationally designed using 2D RNA folding models. The circular RNAs on the other hand were simply made by incorporating the hybridisation region of the previously designed riboregulators within the PIE ribozyme. One major disadvantage of circular RNA parts is that they cannot be easily incorporated within these computational design pipelines using current RNA folding algorithms. This is due to two related reasons. The first is that RNA co-folding software actually calculates the MFE structure of a single strand of RNA,

made by concatenating the two RNA sequences, and removing some of the constraints at the junction of the two RNAs. This is in turn due to the second reason, which is that RNA folding software assumes the absence of pseudoknots (see section 1.0.3). As circularity of one of the RNAs implies the formation of some kind of pseudoknot for interaction, they cannot be handled by current software.

2.3.7 Advantages of circRNA for synthetic Biology

One interesting point is that circular RNA causes a constraint on folding (Lasda and Parker, 2014). Forcing the 5' and 3' ends to be close to each other could force proper folding of an RNA structure that would not fold on a linear strand. This circularity constraint may be responsible for the reduced activity of some of the circular riboregulators described here, as it may interfere with the formation of a stable taRNA-cr-mRNA complex. It could nonetheless be beneficial in other contexts, and further work is required to understand how these constraints influence the interaction of circRNA regulators with their targets. Asides from the introduction of topological constraints, the most notable difference between linear and circular RNAs that could be of use for synthetic biologists is a difference in stability. Circular RNAs have been shown to be more stable than linear ones in bacteria like *E. coli* (Mackie, 2000), as well as in humans (Enuka et al., 2015), presumably due to their resistance to exonucleolytic degradation (Jeck and Sharpless, 2014; Chen, 2016b). This may not be so advantageous for riboregulator circuits in this context of fast growing cells, were the RNAs are expressed from strong promoters, which could allow both linear and circular RNAs to saturate the system. It could however be useful in other contexts. For example, if circRNAs are indeed more stable, they might be used as slow responding regulators, to dampen transcriptional outbursts from other regulators, similarly to their presumed role as 'RNA sponges' in human cells. The increased thermal stability of at least some circRNAs (Abe et al., 2007) may also be useful for using RNA circuits *in vitro* at high temperature, or in thermophilic organisms (Danan et al., 2012), provided a ribozyme

that can function at high temperatures is used.

2.3.8 Using circRNA in other organisms

Whether PIE circularisation would work in thermophiles remains to be seen. However, the autocatalytic nature of group I introns is propitious to their portability across organisms. PIE ribozymes have been shown to work in yeast (Ford and Ares, 1994), and group I introns exist in many other organisms. The possibility that they work in humans would allow potential medical uses. Although PIE ribozymes have rarely been used and have not been shown to work in organisms other than *E. coli* and *S. cerevisiae*, other group I self-splicing introns derived from *T. thermophila* are functional in human cells (Watanabe and Sullenger, 2000; Fiskaa and Birgisdottir, 2010a), which is an encouraging sign that perhaps PIE-circularised synthetic circRNAs could also be made.

2.3.9 Conclusion

Overall, we found that circular riboregulators can be produced by the PIE group I ribozyme method. This method is generalisable to many regulators, and can be used to control fluorescence or antibiotic resistance genes. Toehold accessibility remains an important factor for circular riboregulator design, but other design rules for cr-mRNA design are not valid when using circular rather than linear regulators. The use of circular molecules did not result in increased activation folds compared to when using linear riboregulators. An added layer of complexity is also brought by the self-splicing step, and by the incompatibility of current RNA folding and RNA design software with the circular topology of these molecules. However, PIE ribozymes are simple to design and allow easy *in vivo* circRNA generation. As our understanding of circRNAs improve, and the contexts in which circularisation is an advantage become clear, circular riboregulator production via PIE ribozyme circularisation could become a useful tool for synthetic biologists.

2.4 Materials and Methods

2.4.1 Sequence design and construction

circRAJ31 was designed based on the the principles described in section 2.2.1, using a T4 bacteriphage *td* gene derived group I PIE ribozyme (Sequence supplied by Umekage and Kikuchi, 2009), and synthesised by Integrated Dna Technologies, Inc (Coralville, USA). It was supplied in the pIDT-Amp vector, and the BsaI site present in the AmpR gene was removed by Quikchange mutagenesis. CircRAJ31^{C873U} was generated by Quikchange site directed mutagenesis PCR according to manufacturer's instructions, using primers tdC873U_sense and tdC873U_antisense, whereas circRAJ31^{Toe} was generated by amplifying the whole vector with primers CircRaj_toe_sense and CircRaj_toe_asense and re-ligating the PCR product by goldengate. taRAJ31min was constructed by amplifying circRAJ31 with wr_cr83 and wr_cr84 and cloning the resulting product in the EcoRI site and PstI sites of the pIDT-Amp plasmid. pT7-circRAJ31 and pT7-circRAJ31^{C873U} were expressed from pSB1C3. The cisRAJ31-GFP plasmid was obtained from Rodrigo et al. (2012). It was amplified with GFP_AddLAA_fw/rv and cisR31-Fw/Rv. The former was goldengate ligated with the annealed LAA_anneal_Fw/Rv and construct cisRAJ31-GFP-LAA. The latter was goldengate ligated with the *cat* amplified from pSB1C3 with GG-CAT-Fw/Rv.

Nusl3, Nusl6 and cisRAG parts were synthesised as Gblocks by IDT and cloned into pSC101-sfGFP-DB by goldengate. circRAG parts were bought as primers (circRAG#S and AS) and annealed, then cloned into the PIE ribozyme present in circular_generic_RFP via goldengate (see appendix A.1).

All riboregulators were produced from high copy plasmids (pMB1 origin of replication). All cis-repressed genes were expressed from the medium copy plasmid pSB4K5, which has a pSC101-E93K origin of replication (Peterson and Phillips, 2008).

2.4.2 RNA production

In vitro

pT7-circRAJ31 and pT7-circRAJ31^{C873U} were used as templates to transcribe RNA *in vitro* using the Transcriptaid T7 high yield transcription kit (Life Tech), according to manufacturer's instructions. Briefly, the plasmid containing circRAJ31 and circRAJ31^{C873U} under control of a T7 promoter were cut with PstI to linearise the plasmid, and incubated in the presence of T7 RNA polymerase for 30 minutes to produce RNA. The RNA was then extracted using acidic (ph 4.5) phenol-chloroform (Ambion) according to manufacturer's instructions and precipitated in ethanol in the presence of 50 µg/mL glycogen (Ambion) and 0.3 M sodium acetate. RNA was treated with DNAse I (NEB), then phenol chloroform extraction repeated.

In vivo

pT7-circRAJ31 and pT7-circRAJ31^{C873U} were transformed into *E. coli* BL21(DE3) cells. Three colonies were grown overnight in LB medium, refreshed by diluting them 1:200 in fresh medium in the presence of 125 µM IPTG and 35 µg/mL chloramphenicol for 4 hours before total RNA extraction. For RNA extraction, 20 mL of cells were spun down, resuspended in 400 µL TE buffer, disrupted with 0.1 µm glass beads in a cell disrupter for 5 minutes, then RNA was extracted using acidic phenol-chlorophorm as described above. Extracted samples were run on a 30% polyacrylamide, 6% urea gel to check for RNA integrity before the next steps.

2.4.3 RT-PCR

Primer-specific RT was used to reverse transcribe RNA using circRT-Rv, using M-MuLV reverse transcriptase (Life tech), according to manufacturer's instructions. RNA was then degraded by adding 1 µL of 4M NaOH to the 20 µL RT reaction, heating to 95°C for 3 minutes, then neutralised with 1 µL of 1M HCl. 1 µL of this reaction was used

as a template for the PCR reaction, which was carried out with GoTaq Green Master mix (Promega), using primers CircRT-Fw and CircRT-Rv (for detecting the circularised product), or CircRT-Lin and CircRT-Rv (for detecting the unspliced product). Products were then run on a 2% agarose gel (2h, 85V) and imaged in a Biometra BioDocAnalyze. The rest of the PCR product was purified with a PCR cleanup kit (Thermo Scientific), and sent for sequencing using circRT-Rv as a primer. The chromatograms shown in figures 2.6 and 2.7 were aligned using SnapGene (GSL Biotech LLC).

2.4.4 Fluorescence measurements

For population fluorescence measurements, single colonies of freshly transformed cells were grown overnight in LB medium. The next morning, cells were refreshed in enriched M9 medium (recipe in appendix A, table A.1) by diluting 5µL of culture in 1 mL M9 (1:200 dilution), grown for 6 hours at 37^{o}C with shaking, then refreshed a second time by inoculating 5µL in 200µL of M9 in a 96 well plate (Greiner Bio-One 96 well plates, black, transparent bottom), and grown in a TECAN Infinite F-500 plate reader fluorometer with shaking (2mm, orbital) at 37^{o}C, with measurements taken every 15 minutes. Measurements taken were OD_{600} and GFP (excitation: 465nm, bandwidth 35nm; emission: 535nm, bandwidth 25nm; gain setting 35). Other TECAN settings: Z-position height (manual): 22000µM; lag time: 0 µs; multiple reads per well type: FilledSquare, $2x2$, with Well border: 1450µM(for $OD_{\lambda=600}$) and 900µM(for GFP). For analysis, technical replicates were averaged, values of background substracted, then points closest to O.D. 0.3 were picked for each biological replicate. Bar charts show the average GFP/$OD_{\lambda=600}$of biological replicates, error bars represent standard deviation between biological replicates. For activation fold calculations in figure 2.16, values of GFP/OD_{600} with both inducers was divided by GFP/OD_{600} values with IPTG alone, for each biological replicate. Standard deviation between biological replicates shown as error bars in all cases.

2.4.5 Antibiotic resistance assays

For assays on plates, appropriate plasmids were co-transformed into *E. coli* DH5α cells and three single colonies were picked and grown overnight in LB. The next morning, samples were diluted 1:1000, grown for 2 hours, serially diluted and plated on agar plates containing ampicillin (100 µg/mL), kanamycin (35 µg/mL), spectinomycin (100 µg/mL), was well as, if appropriate, chloramphenicol (35 µg/mL), IPTG (250 µM) and/or aTc (100 ng/mL). Plates were then incubated for 20 hours at 37 °C, colonies were counted and photos taken. For tests in plates, cells were grown overnight, refreshed in M9 medium without inducers, then inoculated into 96 well plates with inducers and chloramphenicol, as appropriate.

2.4.6 Computational design of parts

Nusl3 and Nusl6 cis-repressed UTRs were designed using the NUPACK design webpage (Zadeh et al., 2011). The structural specifications described in figure 2.13 were converted into DU+ notation and automatically designed by the NUPACK webserver. Options used were: Material : RNA; Temperature : 37°C; 10 trials, sodium: 1M, dangles: some. Prevented patterns: AAAA, CCCC, GGGG, UUUU, KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY. The resulting sequences were manually curated, and sequences with stop codons in the fusion protein coding region were discarded. The NUPACK code used for design is given in appendix E.1.

3 Directed evolution of RNA parts with bacteriophage

3.1 Introduction

3.1.1 Directed evolution of biological parts

Directed evolution is another method for the engineering of biological parts. In contrast to rational or computational design, directed evolution relies on the random modification of DNA or RNA sequences, followed by screening or selection of the best parts or organisms, through a user-defined process that aims to imitate natural selection (Lutz, 2011; Packer and Liu, 2015). As this screening or selection process picks the best-working parts amongst a library, there is usually a need to have a working part as a starting point, as non-working parts cannot be selected for. If this is not the case, a working part needs to be created *de novo* during the first round of library generation. Computational design, on the other hand, can be used to create simple working parts *de novo*. However the models used in computational design are necessarily simplified and fail to take into account biological complexity. The parts created may therefore not be optimised for *in vivo* use. Computational design and directed evolution can therefore be seen as complementary techniques. The former can be used for *de novo* design of parts, whereas the latter use these as a starting point for further improvement. As directed evolution aims to blindly select the best

working parts, it can be used for improving parts whose mechanism is not completely understood, or whose activity cannot be predicted computationally. In the realm of RNA part engineering, this is the case for ribozymes or other RNAs where tertiary structure plays an important role. The major disadvantage of directed evolution is the difficulty in designing a selection process that will favour mutants with the desired characteristics, rather than "cheaters" that can outcompete other mutants in an unexpected or undesired way (Raman et al., 2014).

3.1.2 Diversity generation and selection or screening

Directed evolution requires two stages, outlined below: a diversity generation step, and screening or selection step.

Diversity generation

During the first step, a library of sequences has to be generated. There are many methods for generating diversity, and this can be done *in vitro* or *in vivo*. This can be done through imprecise replication of a DNA or RNA template, or through the direct cloning of a synthetic library (Tee and Wong, 2013). Errors could be introduced during replication by the DNA polymerase, but natural mutation rates are often too low to be useful in an artificial setting (Badran et al., 2016), so chemical or physical mutagens are often used to increase this rate. N-methyl-N9-nitro-N-nitrosoguanidine (NG) or ultraviolet light are common examples (Wong et al., 2006; Labrou, 2010).

Due to the risk involved with the manipulation of these chemicals, biological methods have been developed to replace them. Again, methods exist for use both *in vitro* and *in vivo*. The use of error-prone DNA polymerases (Söte et al., 2011), or error-prone PCR conditions if used *in vitro*, can generate diversity during DNA replication. Another method is to decrease replication fidelity by using hypermutator strains (Greener et al., 1997), or mutagenesis-increasing plasmids (Badran and Liu, 2015).

Rather than causing errors during DNA replication, a library of DNA parts can be

prepared synthetically and introduced through cloning and transformation (Tee and Wong, 2013) or *in vivo* recombination with the target (Swers et al., 2004). These library approaches allow targeted mutagenesis, concentrating diversity generation in one or a few "mutational hotspots" that are known (or presumed) to be important for part functionality. This approach has been termed 'semi-rational' engineering (Lutz, 2011).

Screening or selection

During the second step, the mutagenised library is subjected to a custom selection or screening process that picks the best performing members of the library. A screening process consists in analysing each member of the library for functionality, and picking the best performing members. The details of the process used depends on the objectives of the directed evolution. The most straightforward screening approaches directly assay for part functionality, such as product yield in the case of an enzyme (Coelho et al., 2013). Such an approach is limited by its throughput. To permit the faster screening of libraries, one solution is to link part functionality with fluorescence. This can be done by using a surrogate molecule whose reaction with the evolved gene results in the production of fluorescence (Ostafe et al., 2014), or by using genetic reporters which produce a fluorescence output that correlates with desired part functionality (Swe et al., 2012). Combining fluorescence or luminescence outputs with Fluorescence-Activated cell sorting (FACS) can allow rapid screening of cultures at the single cell level, up to 10^8 per day (Packer and Liu, 2015).

By definition, any screening method requires the interrogation of individual colonies or cells, which limits throughput and constrains the methodology used. Selection methods are designed to bypass the necessity to test each phenotype individually and instead links the activity of the part of interest to the reproduction or reproduction rate of the genotype that encodes it. Again, selection can take on different forms depending on the specifics of the activity that is being selected for. One approach consists in selecting good phenotypes based on binding affinity for a target. Proteins

are displayed at the surface of a cell (bacterial display – Bessette et al., 2004), or of a bacteriophage capsid (phage display – McCafferty et al., 1990). Weak binders are washed away, whereas strong binders can be recovered and reused in other rounds of diversification and selection. Binding selection methods have given good results, notably for antibody production, but remain limited in scope. Another selection strategy is to link the survival of the organism to a desired protein function. A simple example is the evolution of expanded spectrum antibiotic resistance proteins, where the evolving sequence directly leads to better growth (Orencia et al., 2001). If the correct function of the part can be linked to gene expression, growth on antibiotics can be used as a proxy to select better parts. For example, Desai and Gallivan (2004) could select riboswitches for their ability to grow in chloramphenicol, and Olson and Müller (2012) selected better trans-splicing ribozymes for their ability to reconstitute a chloramphenicol resistance mRNA. Similar strategies can be devised with auxotroph complementation (Young et al., 2014). In all cases, in order to select for useful parts, the selection methodology must link higher growth rates or survival with activity of the evolved protein, either directly or indirectly through a genetic circuit.

Screening methodologies are usually limited by the throughput of the screening process, with the fastest methods allowing 10^8 test per day. Selection procedures are often limited by the transformation efficiency, which can reach 10^{10} transformants per experiment with *E. coli*. In both cases, the replication and selection rounds are usually iterated until parts with the desired qualities are obtained.

3.1.3 Iterative and continuous directed evolution

Iterative evolution

Traditional directed evolution techniques rely on discrete rounds of genetic variation generation, and rounds of selection. Diversity can be generated by various methods described in section 3.1.2, and by Packer and Liu (2015). This diversity generation allows the creation of a library of parts. It takes place independently in time or space

of the selection process that follows it, during which the best parts are selected for.

Continuous evolution and PACE

Continuous evolution, on the other hand, is the evolution of organisms in continuous culture, where the diversity generation and the selection procedures are taking place concurrently, without any spacial or temporal separation. Evolution experiments in batch culture have been used for long term evolution of bacteria, most notably in the *E. coli* long-term evolution experiment (Travisano et al., 1995; Tenaillon et al., 2016). In this case, cultures of *E. coli* are grown in batch and have been rediluted every day, since 1988. Phage can also be evolved by serial transfer, but as this involves co-culture of bacteria and phage, both organisms tend to co-evolve. Mutations that render the bacteria resistant to the phage can be selected for after a single challenge by bacteriophage (Qimron et al., 2006; Moons et al., 2013). To study phage evolution independently of its host, Husimi et al. (1982) invented the 'cellstat'. A cellstat is a continuously diluted bioreactor which always contains the same number of cells – as opposed to a chemostat or a turbidostat, which maintain at a constant level either the concentration of chemicals (usually nutrients), or the turbidity of the culture (via a feedback between an OD sensor and the dilution rate). The cellstat is maintained by using two interconnected bioreactors. The first is a turbidostat that maintains a constant culture of bacterial cells, at a constant OD. It is connected to the second reactor (the cellstat), where the bacteriophage is grown and which is supplied with a constant influx of bacterial cells. The system is set up such that the average time spent in the cellstat is inferior to the generation time of the bacteria. This ensures that the bacteria do not grow in the cellstat, and therefore not evolve resistance to the phage. As the phage growth rate is much higher than that of the bacteria, phage can sustain their population in the cellstat without any co-evolution of the bacteria.

One recent improvement in continuous directed evolution technology is the development of Phage Assisted Continuous Evolution (PACE) (Esvelt et al., 2011). The PACE

methodology reused this 'evolution machine' devised by Husimi et al. (1982) to evolve bacteriophage, bringing in a key change: One of the bacteriophage genes is removed from its genome, and placed on a plasmid in the bacterial cells (see fig. 3.1). Then, a gene of interest is placed on the bacteriophage genome, in such a way that it controls the expression of the bacterial gene from the plasmid. This way, the gene on the bacteriophage genome evolves to increase expression of the selection gene placed on the plasmid. PACE was initially used to evolve variants of the T7 RNA polymerase that recognise different promoter sequences (Esvelt et al., 2011). The technique has been used to evolve proteases (Dickinson et al., 2014), nucleases (Hubbard et al., 2015) and protein-protein interactions (Badran et al., 2016). PACE has also been expanded to incorporate negative selection, by using a truncated selection gene which is deleterious to the phage when expressed (Carlson et al., 2014), and the authors have worked to increase the mutation rate of the phage by using an inducible mutagenesis plasmid (Badran and Liu, 2015).

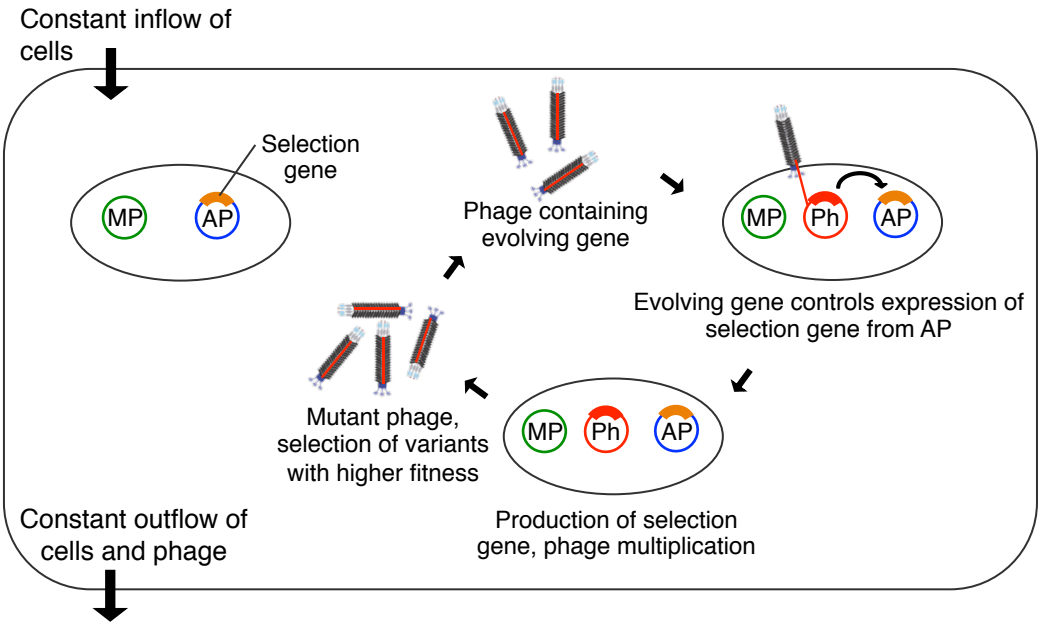


Figure 3.1 – PACE uses a turbodistat to supply exponential phage bacteria to a cellstat which contains continuously replicating bacteriophage. An evolving gene is expressed from the Phage genome (Ph). This gene controls expression of the selection gene from the accessory plasmid (AP). Expression of the selection gene increases production of infectious phage. Mutants that increase expression of the selection gene are therefore selected for. Mutagenesis is continuously provided by the mutagenesis plasmid (MP). Figure adapted from Esvelt et al. (2011).

3.1.4 T7 bacteriophage

T7 bacteriophage is an efficient bacterial killing machine. It is fast-replicating (~17 minutes), lytic, and carries its own transcription and replication systems, making it largely independent from the host machinery (Molineux, 2006). For these reasons, it could be an attractive chassis for Phage Assisted Continuous Evolution (PACE) or PACE-inspired techniques that involve continuous phage growth. Fast replication mean more rounds per time unit, its lytic nature reduce or eliminate biofilm which complicate bioreactor operation, and the independent replication machinery might allow the use of T7 in other hosts. Unfortunately a major drawback to its use is the absence of an easy, robust cloning method for engineering the T7 genome. However, homologous recombination followed by selection has been used to delete portions

of the T7 genome. Selection for mutants is achieved by incorporating the *E. coli* genes *cmk* or *trxA*, that are important for T7 replication, between homology arms. Recombinant phage is then selected by plating on *E. coli* deficient in such genes (Qimron et al., 2006; Tridgett, 2015).

T7 bacteriophage life cycle

The T7 genome is 39,937 bp long, coding for 56 genes (Dunn et al., 1983), encapsulated in a 60 nm icosahedral capsid, with a 23 nm tail (Rontó et al., 1983; Molineux, 2006). T7 can infect *E. coli* as well as some *Salmonella* and *Shigella* strains, and some mutants can infect *Y. pestis* strains (Molineux, 2006). T7 infection starts when the tail fibre interacts with the outer membrane lipopolysaccharide (LPS), the first step of phage adsorption. After binding, the inner core of the phage is ejected from the virion capsid into the outer membrane and periplasm, where it degrades the peptidoglycan layer, presumably forming a channel for DNA translocation, although this has not been well characterised (Molineux, 2006). The first ~850 bp of the genome then enter the cell, allowing initiation of transcription by the *E. coli* RNA polymerase that recognises the promoters present at the start of the genome. Transcription of the genome by the endogenous polymerases pull the rest of the phage genome into the cell (Garcia and Molineux, 1995). This gradual entry of the genome into the host cell forms a first layer of gene regulation. Genes transcribed from the phage can be divided into three classes (see fig. 3.2). Class I genes are expressed early in the infection stage – up to 8 minutes – from three strong *E. coli* promoters (Moak and Molineux, 2000) situated at the start of the genome. They consist of 9 genes, from gene 0.3 to 1.3, expressed from mRNAs transcribed from different promoters but that terminate at the TE terminator that follows *gp1.3*. Two important proteins are gene 0.7 and gene 1. The product of *gp0.7* inactivates host transcription and phosphorylates many host proteins, including ribosomal subunits, RNase III and RNase E, and *E. coli* polymerase subunits (McAllister and Barrett, 1977; Molineux, 2006). Gene 1 codes for T7 RNA polymerase, which goes on to initiate transcription of late genes

from T7 promoters. Late genes are divided into class II and class III. Class II genes are expressed from 6 to 15 minutes after infection at 30°C, and consist of enzymes required for T7 replication. *gp2* also inhibits the *E. coli* RNA polymerase (Nechaev and Severinov, 1999), switching off transcription of early genes. *gp3* and *gp6* are responsible for degrading the host chromosome. These nucleotides are then recycled by endogenous *E. coli* proteins, and make up 80% of the progeny phage DNA. The Cmk protein is responsible for the recycling of cytidine monophosphate (CMP) nucleotides. Its deletion greatly reduces phage replication, allowing its use as a selection gene. *gp5* is the T7 DNA polymerase which replicates the phage chromosome, in complex with the *E. coli* protein thioredoxin (TrxA). TrxA acts as a processivity factor for the phage DNA polymerase, increasing the speed of DNA polymerization 1000 fold, and is essential for T7 replication. Class III genes are responsible for DNA packaging, viron assembly and the eventual lysis of the host which releases the virons (Molineux, 2006). Class III gene mRNAs are highly expressed, which may reduce class II gene expression by ribosome sequestering.

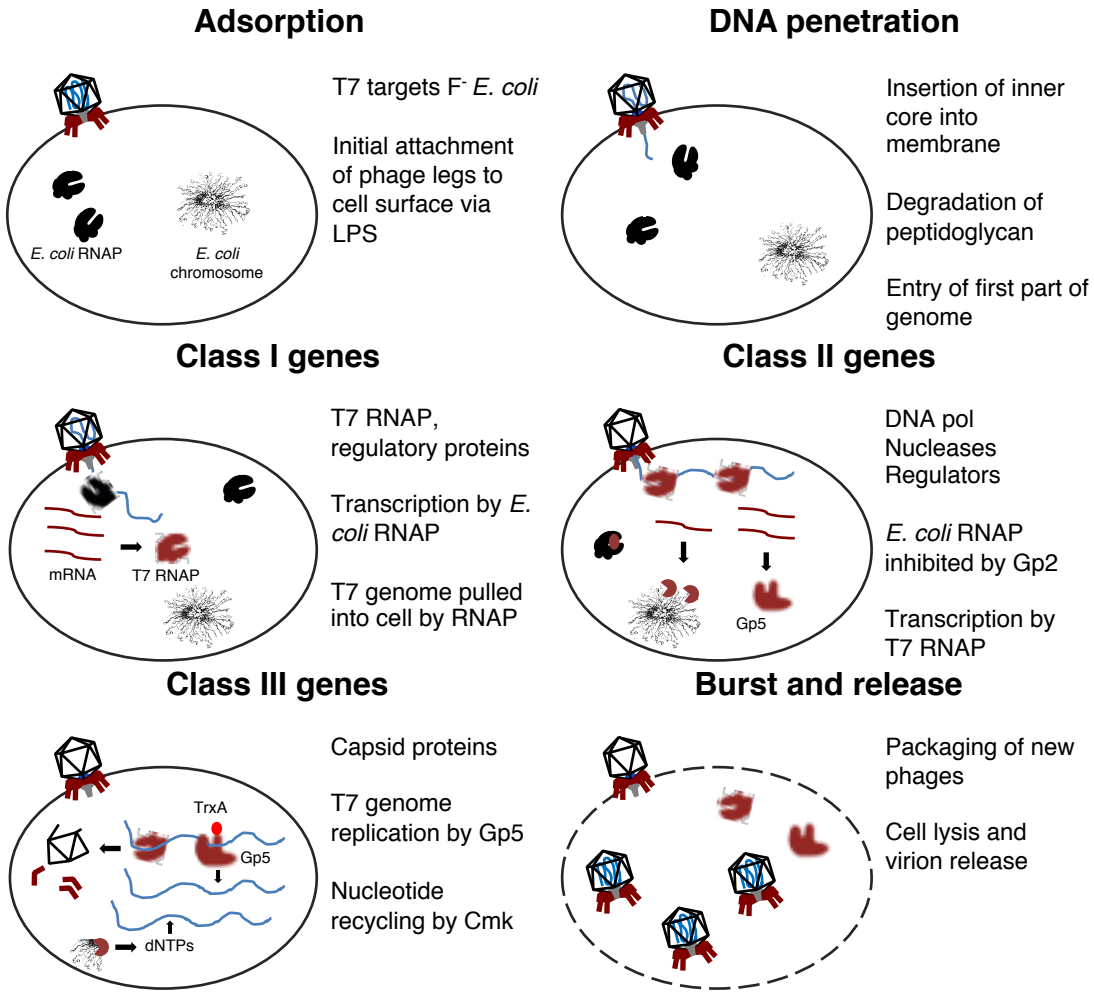


Figure 3.2 – The life cycle of T7. Adsorption starts with attachment of phage legs to LPS. The inner core enters the cell membrane, 850 bp of the genome enter the cytoplasm, where the *E. coli* RNAP expresses Class I genes, simultaneously pulling in the rest of the genome. T7 RNAP expresses Class II then III genes. The T7 genome is replicated by the Gp5-TrxA complex. Nucleases degrade the *E. coli* genome, and nucleotides are recycled.

3.1.5 Engineering T7

Even though bacteriophage have recently attracted much attention as a potential alternative to antibiotics, engineering of lytic phage like T7 remains relatively rare. Modifications are complicated by the fact that T7 is a lytic phage with a dsDNA linear genome. Unlike M13 or other lysogenic DNA phages with circular genomes, it cannot be transformed into a phagemid – a hybrid that contains both a phage and a plasmid origin of replication. There are nonetheless a few methods to engineer T7. One approach uses homologous recombination, either in combination with a T7 selection marker, or by targeting and destroying non-recombinant phage with CRISPR (Qimron et al., 2006; Kiro et al., 2014). In both cases, the exogenous gene is placed between homology arms on a plasmid in *E. coli* cells, which are then killed with wild type T7. If the homologous recombination cassette includes a selection marker, the lysate is plated on *E. coli* cells where this gene is deleted. Two previously used selection markers are thioredoxin (*trxA*) and cytidine monophosphate kinase (*cmk*), which can be inserted onto the T7 genome and selected for using Δ*trxA* or Δ*cmk E. coli* cells (Qimron et al., 2006). Both genes are non-essential for *E. coli* but essential (in the case of *trxA*) or very important (for *cmk*, without which efficiency of plating of T7 is reduced by 95%).

Instead of using a selection gene, nucleases can be targeted to the wild type sequence, favouring recombinant bacteriophage. CRISPR-Cas9 has been used in this way to engineer T7 (Kiro et al., 2014). Alternatively, the whole T7 genome can also be transferred into a Yeast Artificial Chromosome (YAC), in one piece or as PCR fragments assembled through gap-repair cloning (Jaschke et al., 2012; Lu et al., 2013). The phage genome can then be modified, before 'rebooting' the phage by transformation into *E. coli*. This method has been used to engineer T7 with modified host ranges (Ando et al., 2015).

3.1.6 Continuous evolution of T7 bacteriophage

Directed evolution could be a good complement to computational or rational design strategies for creating new genetic parts. Amongst directed evolution techniques, continuous evolution is a promising method for the improvement of biological parts. However, it has so far only been carried out with M13, even though other organisms such as T7 could bring advantages, as they are lytic and can replicate in other hosts. This chapter presents an investigation of whether T7 can be used in continuous evolution experiments to optimise previously designed parts. First, methods for creating engineered T7 containing exogeneous cargo genes are tested using two previously used selection markers. The validity of these recombination techniques is tested in batch culture, but also in continuous culture within continuously diluted bioreactors. The effectivity of two potential selection genes, the T7 DNA polymerase *gp5* and the *E. coli* gene *cmk*, are tested. A recombinant T7 phage containing the circRAJ31 riboregulator was then grown in a cellstat in continuous culture, and the accumulation of mutations in the phage in presence of an error prone DNA polymerase measured.

3.2 Results

3.2.1 Construction of recombinant T7 bacteriophage

To construct recombinant T7 bacteriophage, initially, both the *cmk* and *trxA* genes were used as selection markers, to increase the chances that any cargo carried between them was transferred to the phage genome (although one selection gene was later found to be sufficient). The selection cassette containing both genes between 100 bp homology arms was transferred into pLitmus-chlor-RFP, generating PR100-*cmk-trxA*. The homology arms contain the last 100 bp of the T7 gene 4.7 and the first 100 bp of the T7 gene 5.3. This targets the recombination cassette to the gene 5 locus (T7 DNA polymerase), removing it in the process. *E. coli* MG1655 cells carrying the recombination plasmid were infected with wild type T7 phage (MOI 0.01), which resulted in bacterial lysis. Dilutions of filter sterilised lysate (containing 200 µL , 20 µL and 2 µL of lysate) were used to challenge Δ*cmk* Δ*trxA E. coli* cells containing pT7-G5. The 200 µL dilution killed all bacterial cells, and the 2 µL dilution resulted in a bacterial lawn with no visible plaques. The intermediate 20 µL dilution resulted in an uneven lawn of small colonies spread throughout the plates, interspersed with a few T7 plaque-sized clearings. To confirm that these clearings were recombinant T7 plaques, the *gp5* region was amplified by PCR with phage specific primers located outside the target site. A product of the expected size (3000 bp) was detected and sequenced, confirming the insertion of the recombination cassette containing *cmk* and *trxA* into the T7 genome, and the removal of *gp5*. This validated this method for recombinant phage generation, and the recombinant phage was named T7:Δ*g5*::*cmk-trxA*. The cI gene from phage λ (λ*cI*) was then inserted between the selection genes on PR100-*cmk-trxA*, generating PR100-*cmk*-λ-*trxA*, and the recombination process repeated. Initially, no recombinant phage were detectable, as the undiluted sample (200 µL of lysate) killed all cells whereas diluted samples (20 µL and 2 µL dilutions), a lawn or interspersed colonies grew, with no visible plaques. The killing of bacterial cells is

due to the presence of wild type phage in the lysate. Wild type T7 is non-replicative in Δ*cmk* Δ*trxA* cells, but can still cause abortive infection and cell death. As a result, there is a theoretical limit to recombinant phage detection which corresponds to the concentration where all cells get killed by wild type T7, preventing the growth of a cell lawn and therefore the detection of recombinant T7 plaques. To detect whether recombinant phage was present at lower levels, an amplification stem was carried out. 150 mL of liquid culture of Δ*cmk* Δ*trxA E. coli* cells was inoculated with 1 mL of lysate, and the new lysate recovered and filtered. Plaque assays were repeated with this 'post-selection' lysate, and recombinant T7:Δ*g5*::*cmk*-λ*cI*-*trxA* were picked after subsequent plaque assays on Δ*cmk* Δ*trxA* cells. The presence of the λcI containing selection cassette on the phage was confirmed by PCR and sequencing, validating this method for generating recombinant T7 bacteriophage.

3.2.2 'On the fly' recombineering with phage co-infection

With the confirmation that exogenous cargo could be inserted onto the T7 genome to create recombinant bacteriophage, the possibility of carrying out recombination during continuous culture of bacteriophage was investigated. The ability to supply exogenous sequences during continuous culture of bacteriophage would allow the integration of library based screening techniques with continuous directed evolution techniques like PACE, as arbitrary sequences could be supplied for recombination during continuous evolution. To test whether such 'on the fly' recombineering is possible, a system where both M13 and T7 can co-infect a cell was tested.

In wild type strains, superinfection by these two phages does not occur. M13 can only infect F+ *E. coli*, but F plasmids confer resistance to T7 through the *pifA* gene (Morrison and Malamy, 1971). To permit co-infection of T7 and M13 phages, an F+ Δ*pifA E. coli* strain was obtained, and its susceptibility to co-infection was tested. The strain was grown and infected with either M13KO7, wild type T7, both or neither. M13KO7 contains a kanamycin resistance gene, allowing detection of M13 infected

Kanamycin	M13	T7	Result
No	No	No	Uniform lawn
	No	Yes	Lawn with plaques
	Yes	No	Uniform lawn
	Yes	Yes	Lawn with plaques
Yes	No	No	No lawn growth
	No	Yes	No lawn growth
	Yes	No	Uniform lawn
	Yes	Yes	Lawn with plaques

Table 3.1 – Growth of Δ*pifA* cells after infection with or without M13 phagemid (with MOI $\gg 1$) and/or T7 phage (10 PFU per plate). All plates contained Ampicillin and Tetracycline.

cells as they grow on kanamycin containing plates, whereas T7 infection creates characteristic plaques. Only samples infected with M13 grew on kanamycin plates. T7 plaques grew on all bacterial lawns that were inoculated with T7, including M13 infected bacterial lawns grown on kanamycin plates (see table 3.1). These results confirm that F+ Δ*pifA E. coli* can be co-infected by both M13 and T7 bacteriophage.

M13 phagemid particles containing the PR100-*cmk*-*trxA* recombination cassette (M13::PR100-*cmk*-*trxA*) were then generated. As a negative control, phagemids containing an RFP gene (M13::RFP) were made. Δ*pifA* cells containing pT7-G5 were co-infected with M13::PR100-*cmk*-*trxA* or M13::RFP, and with wild type T7. T7 infection of cells previously transduced and selected for PR100-*cmk*-*trxA* uptake was carried out alongside as a positive control. The infected cultures cleared after two hours. To assay the presence of recombinant T7 phages containing the cassette provided by M13 phagemids, the filtered lysates were used to perform plaque assays on *E. coli* Δ*cmk* Δ*trxA* , with or without an amplification step. In non-amplified samples, recombinant phage were detected in pre-transduced *E. coli* lysates, but not in lysates resulting from co-infection. After amplification, recombinant phage was detected in lysates from co-infection with PR100-*cmk*-*trxA* phagemid, but not co-infection with RFP phagemid (see fig. 3.4). This confirmed that recombination between co-infecting M13 phagemid and T7 can take place *in vivo*. The inability to detect recombinant

phage in non-amplified samples shows that recombination after co-infection occurs at a lower level than in previously transformed or transduced cells.

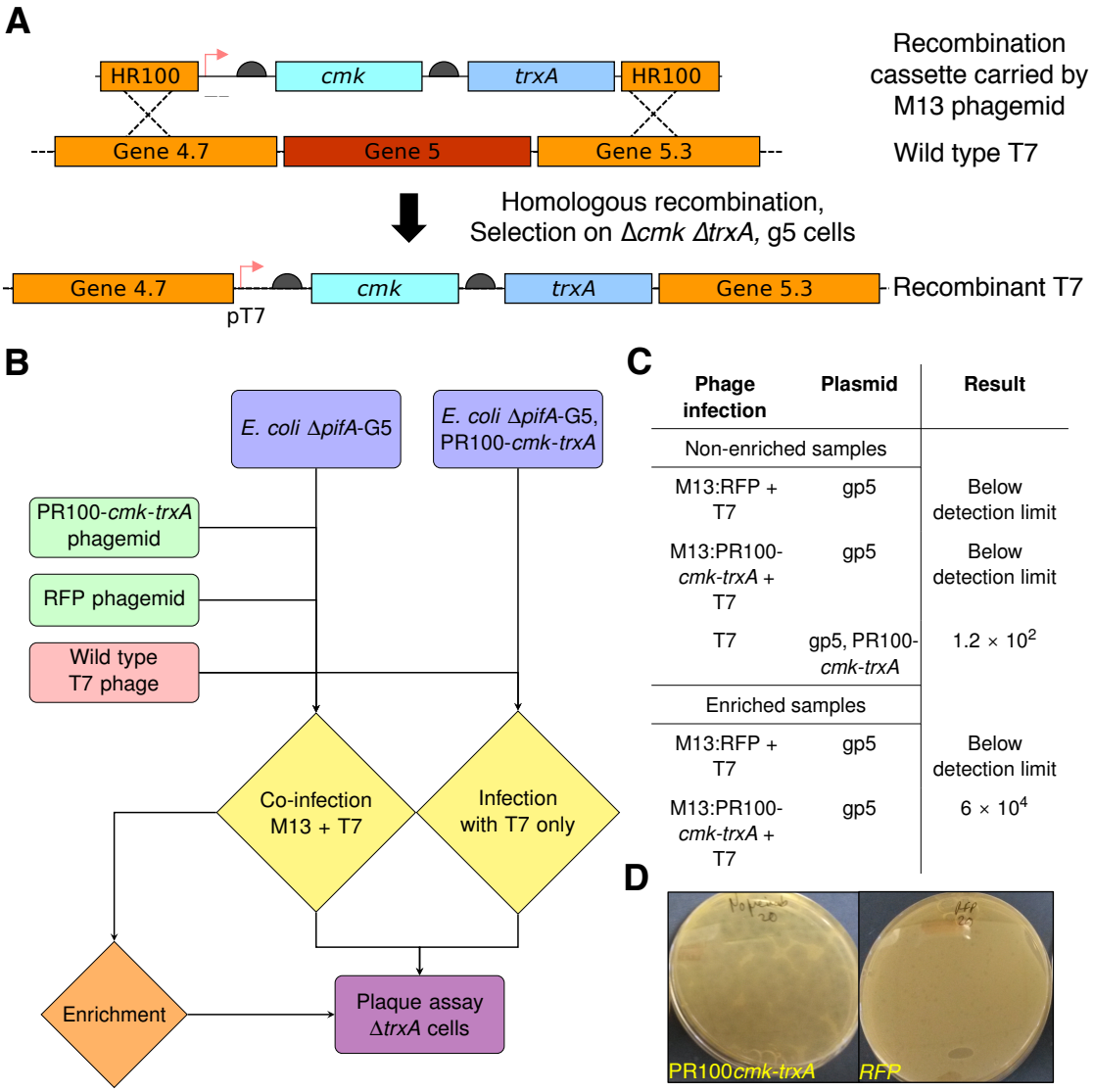
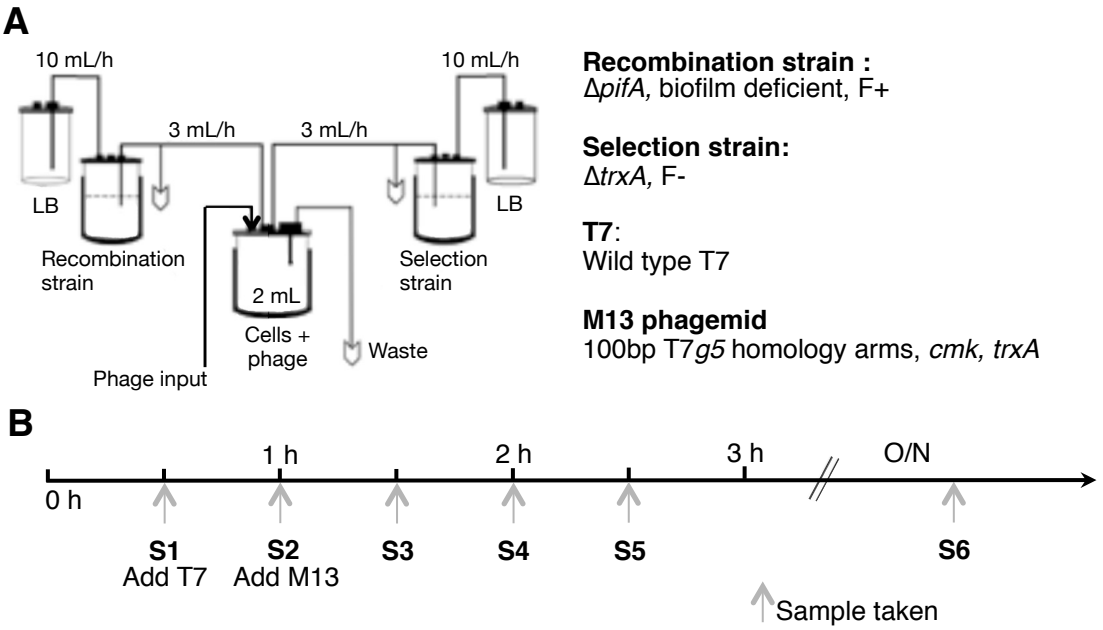


Figure 3.3 – **A**. Schematic of homologous recombination for engineering of T7 bacteriophage. The recombination cassette is brought by an M13 phagemid. *trxA* is essential for phage growth, and recombinant phage are selected for on Δ*trxA* cells. **B**. Diagram of co-infection experiment. T7 and M13 phagemid are used to co-infect *E. coli* cells. **C**. Table of results after co-infection. Recombination after co-infection with M13 phagemids is detectable after sample enrichment. **D**. *E. coli* lawn infected with enriched lysates from co-infections with wild type T7 and M13:PR100*cmk-trxA* (left, plaques visible) or M13:RFP (right, no plaques).

To test the validity of this method alongside continuous evolution, the process of infection was repeated in a continuous T7 culture using interconnected chemostats and cellstats. Two cultures, one containing Δ*pifA* -g5 cells (the recombination strain), and the other a Δ*cmk* Δ*trxA* culture (the selection strain), were grown in chemostats. Both chemostats were connected to a single continuously diluted cellstat in a 50/50 ratio (see fig. 3.4), creating an environment where phage could grow in a mixed bacterial culture. The cellstat was first inoculated with wild type T7, then with M13::PR100-*cmk-trxA*, with samples taken from the cellstat output every hour. The last sample was taken after infection of the chemostats with the M13 phagemids. The presence of recombinant phage in the cellstat samples was then assessed by lysis assays on Δ*cmk* Δ*trxA* cells. In samples 1 (before phage addition), 2 (before M13 phagemid addition), 3 and 4 (after M13 addition in cellstat), no recombinant phage was detected. Sample 5 (after M13 phagemid addition in chemostat) contained recombinant phage. Recombination did not occur if co-infection took place in the cellstat, but did work effectively if M13 was added to the cells in the chemostat, which gives more time for phagemid entry into the cells.



C

Sample	Lyses Δ*pifA* F+ culture	Lyses Δ*cmk* Δ*trxA* culture
1	No	No
2	Yes	No
3	Yes	No
4	Yes	No
5	Yes	No
6	Yes	Yes

Figure 3.4 – **A**. Bioreactor setup, with recombination strain cells grown in one chemostat, selection strain cells grown in another. Both chemostats feed a cellstat where T7 is grown continuously, and M13 is injected intermittently. **B**. Timeline of experiment. **C**. Characterisation of cellstat output. Lysis of Δ*pifA* cells indicates T7 presence, lysis of Δ*cmk* Δ*trxA* cells indicates recombinant T7 presence. Co-infection in cellstat does not permit recombination (Samples 2-5), but M13 addition in chemostat does (sample 6).

3.2.3 Accessory plasmid testing for T7 PACE

Conducting PACE experiments using T7 bacteriophage requires a selection gene whose expression is correlated with higher and faster production of infective offspring. Two genes, *gp5* (T7 DNA polymerase) and *cmk* (cytidilate monophosphate kinase), were tested for their ability to be used as selection genes.

Testing *gp5* as a potential selection gene

To determine whether *gp5* can be used as a selection gene, the ability of two phages, T7ΔG5::*cmk-trxA* and T7ΔG5::*cmk-λcI-trxA* (described in section 3.2.1), to infect *E. coli* cells containing *gp5* on accessory plasmids (AP) was assayed. Two phages (with or without a cargo) were used to determine whether the presence of the cargo influenced phage infectivity, and three APs were used to determine whether the promoter controlling *gp5* transcription influenced phage infectivity. Both phages were diluted to identical concentrations, then used to perform plaque assays on *E. coli* MG1655 cells containing the APs. The phage gene 1 (T7 RNApol), the cargo (λ*cI*) and the gene 5 present on the APs form genetic circuits that all lead to the production of Gp5 (T7 DNA pol) and therefore phage replication (see fig. 3.5A). The three APs all contain *gp5*, downstream of either a pT7 promoter, a pRM promoter with the wild type RBS, or pRM with a weak RBS (B0031). The T7 polymerase (gene 1) is initially transcribed by an endogenous *E. coli* polymerase. T7 RNAP can then transcribe the synthetic unit which is inserted in the gene 5 locus. In the control phage 1, this contains *cmk* and *trxA*, whereas in phage 2, it also contains λcI. In cells containing AP1, T7RNApol directly transcribes G5, whereas in cells containing AP2 and AP3, G5 is under the control of the λcI activated pRM promoter.

The presence of cI on the phage did not lead to higher plaque numbers when cells contained a pRM promoter upstream of Gp5. Infection by both phages was similar, regardless of the accessory plasmid used. The presence of the cI activator gene on the phage DNA in combination with its target pRM promoter on the AP did not lead to

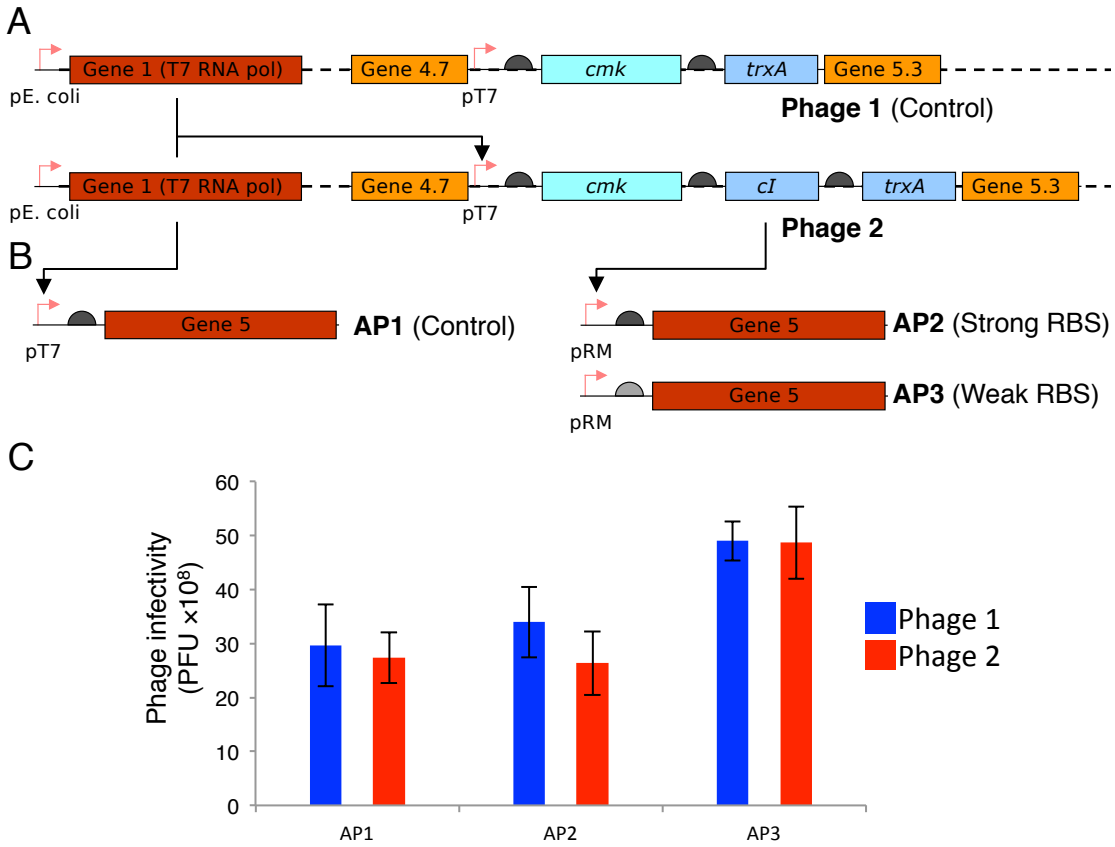


Figure 3.5 – **A**: Recombinant phage containing either *cmk* or *cmk*-λ*cI*-*trxA* inserted into the gene 5 locus. **B**. Accessory plasmids (AP) used. AP1 contains pT7-*g5*. AP2 and AP3 contain pRM-*g5*, with the wild type RBS from AP1 or B0031, respectively. **C**. Results of the plaque assays. Both phage were diluted to identical concentrations beforehand (measured on cells containing AP1), then against MG1655 cells containing each AP.

changes in phage infectivity compared to infection of the same cells with a control phage which does not contain the cI gene. Although we cannot discard the possibility that cI is not expressed from the phage, plaque forming unit (PFU) measurements performed with the control phage were similar when plaque assays are performed on cells expressing *gp5* from a pRM promoter and on cells expressing *gp5* from a strong T7 promoter. This showed that pRM is a leaky promoter which that does not require the presence of cI to be expressed. Varying the AP did cause small differences in plaque numbers. In cells containing AP3, the PFU titres of both phages were slightly higher than in cells containing either AP1 and AP2. This difference was visible regardless of

the phage used. The B0031 RBS was chosen for AP3 because it had a small predicted Translation Initiation Rate (TIR) of 972.90 according to the Sallis lab RBS calculator (Salis et al., 2009; Espah Borujeni et al., 2014). This is much weaker than the wild type RBS used in AP1 and AP2, which has a predicted TIR of 38695.20. Even if the increase in PFU is small on cells containing AP3, these results suggest that *gp5* is not a suitable selection gene for conducting directed evolution with T7. A good selection gene would cause clear differences in infectivity and plaque morphology, which were not detectable when varying the promoter in front of *gp5*. Rather, the leakiness of the pRM promoter without cI presence was sufficient to give the same results as the positive control with a pT7 promoter, and a predicted lower expressing RBS had higher levels, suggesting that lower levels of GP5 could in fact be advantageous.

Testing *cmk* as a selection gene

Cytidine monophosphte kinase (*cmk*) catalyses the phosphorylation of CMP and dCMP to CDP and dCDP, respectively, and is a key part of the salvage pathway that recycles pyrimidine bases (see fig. 3.6A). *cmk* has been shown to be important for T7 growth (Qimron et al., 2006). When T7 enters its host, nucleases are expressed and degrade the *E. coli* genome. The nucleotides that result from this degradation are reused for T7 genome replication. *cmk* deletion results in a 95% reduction in efficiency of plating, and it can be used as a selection marker for T7 recombineering (Qimron et al., 2006). If restoring its activity results in an increase in phage fitness, *cmk* could be used for selection during directed evolution (see fig. 3.6B). To test whether *cmk* could be used as a suitable selection gene on an accessory plasmid, a new recombinant T7 phage containing *trxA* (but not *cmk*) in the gene 5 locus was constructed (see fig 3.7A).

The *trxA* gene was found to be sufficient for recombinant phage selection, and T7Δ*g5*::*trxA* was generated by homologous recombination as previously described. To determine the effect of *cmk* on the phage, phage challenges of *E. coli* cells were performed on plates then in liquid culture. First plaque assays of T7Δ*g5*::*trxA* on either

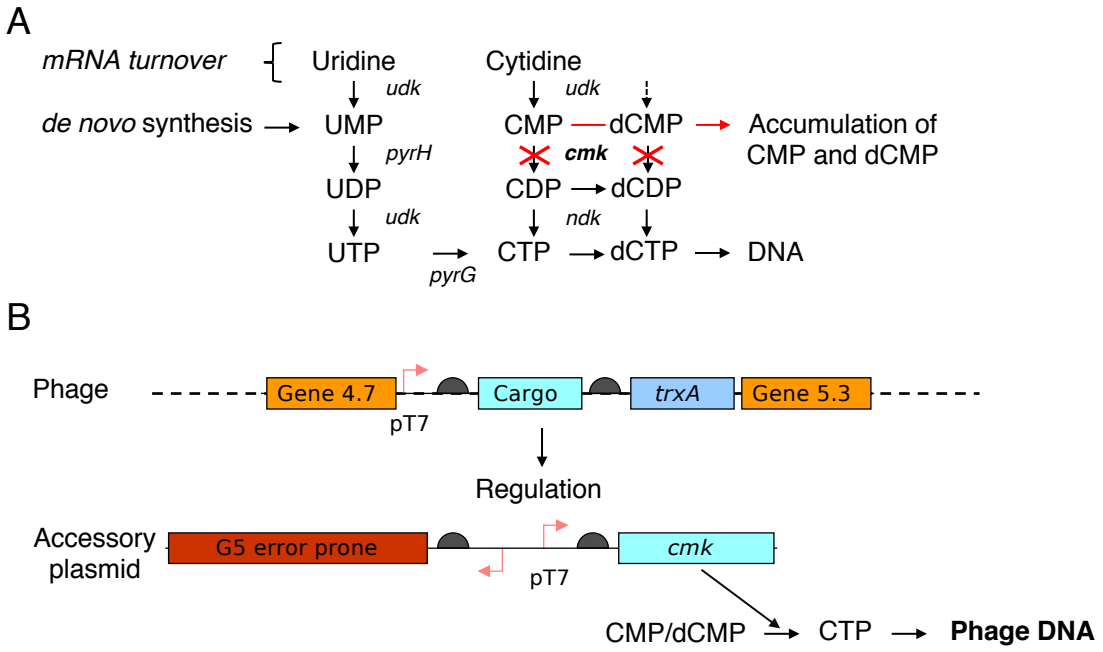


Figure 3.6 – **A**. Pathway to dCTP production and effect of *cmk* deletion. *cmk* is essential for efficient T7 replication. **B**. Putative pathway for phage production via *cmk* production circuit. The phage contains the cargo added using *trxA* as a selection gene. The cargo controls the gene circuit of interest, that leads to *cmk* production.

*E. coli*Δ*trxA* or *E. coli*Δ*cmk* Δ*trxA* cells expressing *gp5* were performed. T7Δ*g5*::*trxA* plaques were found on both *E. coli* strains, but plaques on Δ*cmk* Δ*trxA* cells were much smaller than plaques on the Δ*trxA* control strain (see fig. 3.7). This shows that *cmk* is not essential, but that *cmk* expression does have a large effect on the phage. As PFUs of Δ*g5* phage on cells containing or not containing the *cmk* gene were similar, but with plaque size differences, it was hypothesised that an equal amount of T7Δ*g5*::*trxA* should kill a population of *E. coli* cells slower when *cmk* levels are lower. A kill curve assay was then devised for comparing the effects of *cmk* expression on phage killing speeds. If cells were equally susceptible to infection but phage replication was slower, or burst size smaller, we would expect a similar PFU measurement, but a slower overall time to kill a population of cells. To test this, 4 new accessory plasmids were constructed. They all contain an error prone version of *gp5*, a T7 DNA polymerase engineered to have a decreased replication fidelity (Söte et al., 2011). They also contain the *cmk* gene, but with varying promoters and cis-repressing 5' UTRs

upstream of *cmk* (see fig. 3.8). AP$_{cmk}$-1 contains a pT7 promoter, and was used as a positive control. AP$_{cmk}$-2 contains no promoter or RBS, but a 14 bp spacer, and was used as a negative control. AP$_{cmk}$-3 and AP$_{cmk}$-4 contain a pT7 promoter and a cis-repressed *cmk* gene where the RBS and ATG are repressed by a 5'UTR that block initiation via base pairing. These 5' UTRs were designed to be activated by the RAJ31 riboregulator described in chapter 2, therefore permitting its improvement by evolution. They were designed using NUPACK, with the same design goals as in figure 2.13 (chapter 2).

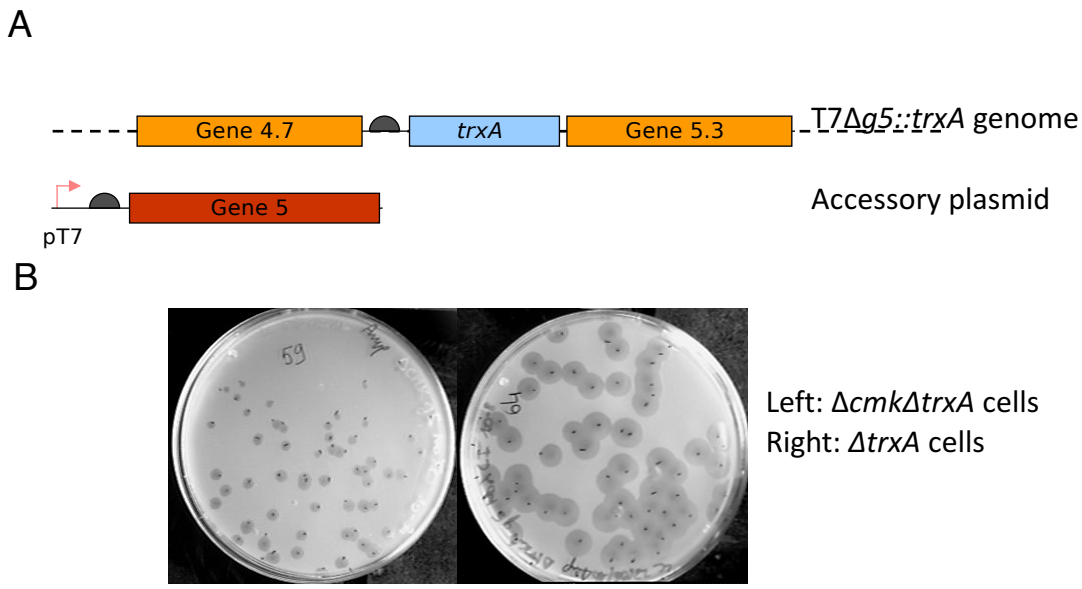


Figure 3.7 – **A**. Contructions used. **B**. Plaque assays with T7ΔG5::*trxA* in Δ*trxA* and Δ*cmk* Δ*trxA E. coli* cells.

The 4 new APs were transformed into *E. coli* Δ*cmk* Δ*trxA*. The cells were grown to OD$_{\lambda=600}$ 0.15, challenged with an equal amount of phage and grown in a plate reader to follow growth and death of the cells. Cells containing pT7-*cmk* reached peak OD$_{\lambda=600}$ after 79 minutes, whereas cells containing no promoter upstream of *cmk* reached peak OD$_{\lambda=600}$ after 102 minutes. Presence of a cis-repressing UTR upstream of the RBS caused slower killing. Cells expressing UTR1 or UTR2 reached peak OD after 94 and 102 minutes, respectively, which is between the peak times for AP $_{cmk}$-1 and AP $_{cmk}$-2. All strains had similar growth rates until they reached their peak OD$_{\lambda=600}$,

apart from cells containing UT2 (AP $_{cmk}$-4) which grew slightly slower. The difference in killing time between the cells expressing AP$_{cmk}$-1 and AP$_{cmk}$-2 indicates that a pT7 controlled *cmk* gene can reduce phage killing times of cells. pT7 is induced by the T7 gene 1 (T7 RNAP), so control of *cmk* expression by another phage gene, whether natural or synthetic, could potentially control *cmk* expression, and therefore phage fitness. The *cmk* gene seemed to be a suitable choice as an accessory selection gene for conducting PACE with T7 bacteriophage, as lower expression increases killing time. The cis-repressing UTRs also slowed the killing of cultures relative to the positive control with only a pT7 promoter.

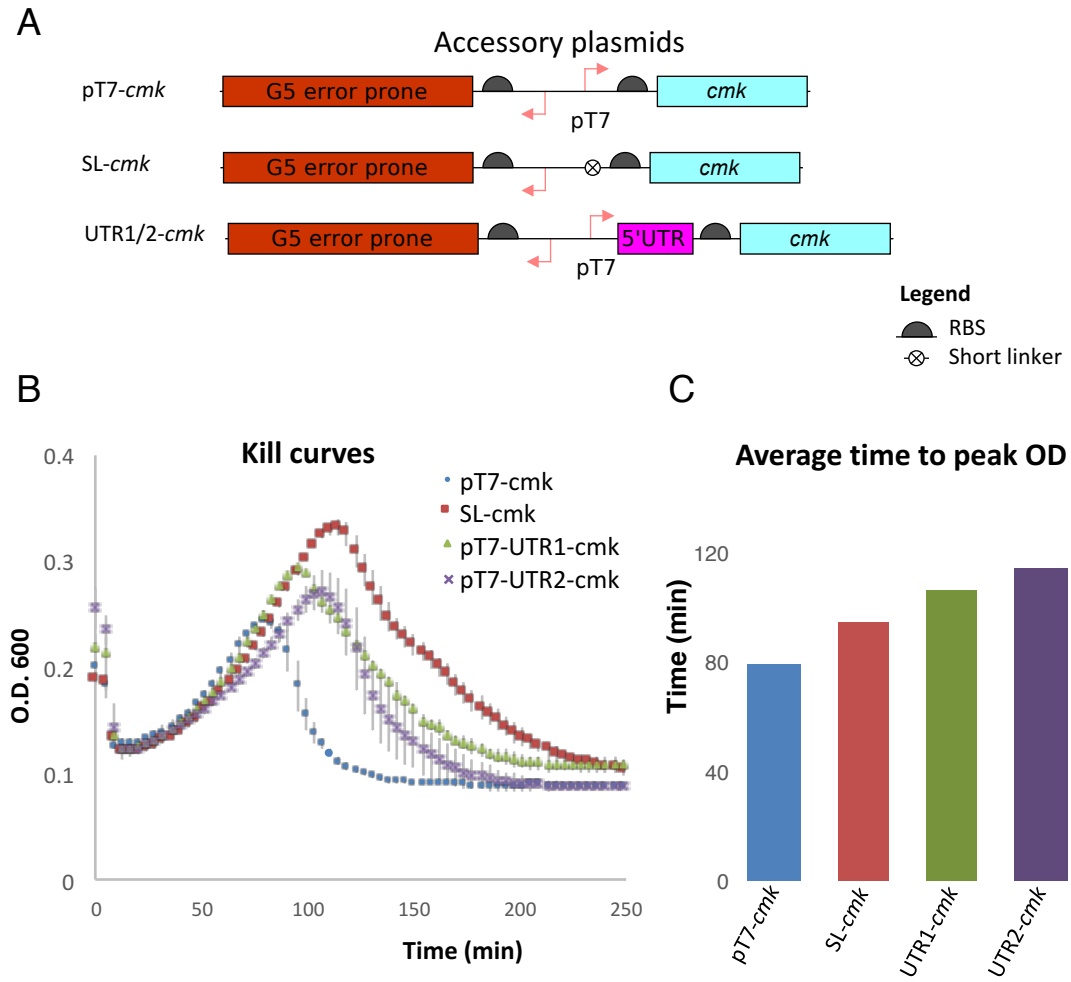
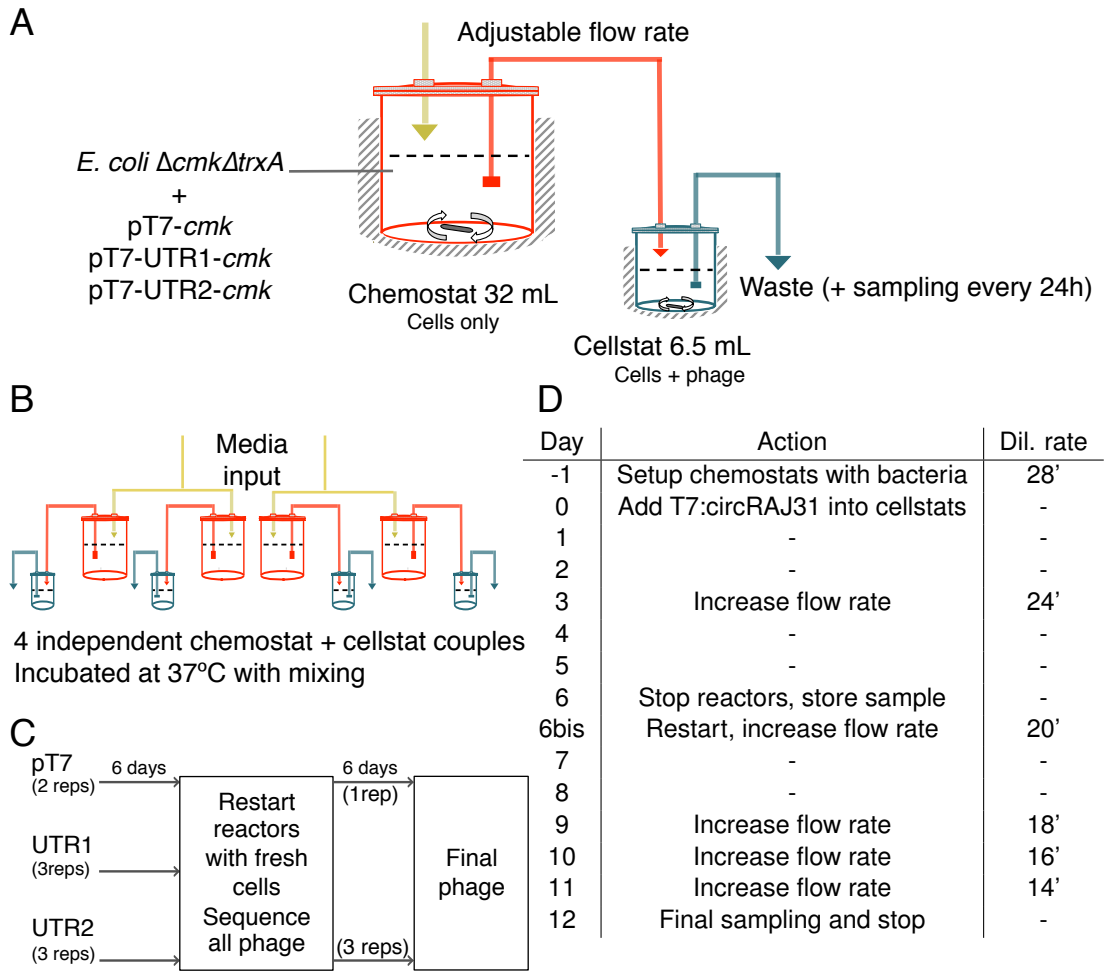


Figure 3.8 – **A**: Accessory plasmids used. pT7-*cmk* and SL-*cmk* (Short linker, no promoter) were used as positive and negative controls. Two cr-mRNA UTRS (UTR1 and UTR2) were tested.**B**: Kill curves of *E. coli* cells containing the different APs described in A, challenged with T7ΔG5::*trxA*. Cultures prepared from three independent colonies were used, and standard deviation of OD measurements is shown as error bars. **C**. Average time to peak $OD_{\lambda=600}$ of cells when infected with equal amounts of phage. Phage fitness inversely correlates with time. Time to peak $OD_{\lambda=600}$ did not vary between replicates, so no error bars are plotted.

3.2.4 Continuous evolution of recombinant T7 bacteriophage

The circular riboregulator RAJ31 (see chapter 2) was then recombineered into T7, producing T7Δ*g5*::pT7-circRAJ31-*trxA*, referred to as T7-circ31 henceforth. A setup of continuously diluted bioreactors were assembled, such that one bioreactor, the chemostat, would continuously maintain a culture of growing, uninfected *E. coli* cells; and be connected to a second bioreactor, the cellstat, inoculated with T7-circ31 (see fig. 3.9 A). The phage can replicate within the continuous supply of cells, maintaining a continuous culture of bacteriophage. The setup is similar to the PACE technique (Esvelt et al., 2011), although it uses T7 phage intead of M13, and *cmk* as a selection gene (instead of M13 gene 3). 4 independent bioreactor setups were run in parallel. The chemostats contained Δ*cmk* Δ*trxA* cells transformed with either pT7-*cmk*, pT7-UTR1-*cmk* or pT7-UTR2-*cmk*, and the cellstats were inoculated at time 0 with T7-circRAJ31 (see fig 3.9 B). The bioreactors were then run for 14 days, starting with a dilution rate of 28 minutes (*ie* a renewal of volume every 28 minutes). The flow rate was increased gradually, to reach a final flow rate of 14 minutes. Samples were taken every 24 hours, and the region containing circRAJ31 was amplified by PCR and sequenced (see fig. 3.9C and D).



Day	Action	Dil. rate
-1	Setup chemostats with bacteria	28’
0	Add T7:circRAJ31 into cellstats	-
1	-	-
2	-	-
3	Increase flow rate	24’
4	-	-
5	-	-
6	Stop reactors, store sample	-
6bis	Restart, increase flow rate	20’
7	-	-
8	-	-
9	Increase flow rate	18’
10	Increase flow rate	16’
11	Increase flow rate	14’
12	Final sampling and stop	-

Figure 3.9 – **A**. Specifications of bioreactors used for evolution of T7-circRAJ31. **B**. Overview of bioreactor setup, with 4 parallel chemostats and cellstats. **C**. Schematics of experiments. 2 replicates of pT7, and 3 replicates of both UTR1 and UTR2 were run for 6 days. 1 replicate of pT7 and all replicates of UTR2 were run for an additional 6 days (total 12 days), before sequencing. **D**. The cellstats were inoculated with T7 on day 0 and the dilution rate was increased gradually during the experiment.

3.2.5 Phage sequencing

Testing for presence of the phage by PCR and/or by plating showed that T7 phage was present in all four cellstats throughout the experiment. Despite the increasing dilution rate, T7 was never washed out. Daily samples were sequenced around the region of the inserted circRAJ31 gene (see fig. 3.10) using Sanger sequencing after PCR of the relevant region. This revealed that all phage populations accumulated different mutations in the bioreactor, in all three genes sequenced (*gp4.7*, *circRAJ31* and *trxA*). Most of the mutations were fixed in the bioreactor after 12 days, with the original genotype undetectable. For some positions, both mutated and non-mutated phage were detectable, showing that some diversity still existed within the phage population in the bioreactor.

3.2.6 Mutations in protein coding genes

Mutations in gene 4.7 were common, with 2.75 mutations in *gp 4.7* on average. There was no apparent pattern, with synonymous, non-synonymous and non-sense mutations appearing in the phage genome (see fig. 3.10B). Several non-synonymous mutations were fixed in samples pT7(2), UTR2(2) and UTR2(3), and a nonsense mutation was fixed in sample UTR2(3). Although it is not known what effect the mis-sense mutations have, the non-sense mutation in sample UTR2(3) truncates Gp4.7 to 42% of its original length. The accumulation of non-synonymous mutations as well as the subsistence of the genotype containing a non-sense mutation in the bioreactor does not show any evidence of selection pressure operating on *gp4.7*, and suggests that deletion of this gene is not deleterious.

13 different mutations appeared in the *trxA* coding sequence in the four bioreactors, an average of 3.25 mutations per sequence in a 333 bp long coding region, although with uneven numbers of mutations in each bioreactor. Three samples contained 1 or 2 mutations, whereas UTR2(1) had 8 mutations after 2 weeks of continuous growth

(see fig. 3.10C). Out of 13 mutations that appeared in the *trxA* gene, 7 were silent mutations (61%), whereas the probability that a random mutation in *trxA* is silent is 24%, assuming an equal probability of any mutation happening. Out of the 6 non-silent mutations, 4 mutated the same amino acid Arg74. These 4 mutations led to 3 amino acid changes: Arg74Cys in Nu2_1 and in Nu2_2, and Arg74Leu in Nu2_3 (the other mutations caused Asp3Asn and Ala57Thr in Nu2_1 and Nu2_2, respectively). The high amount of synonymous mutations within the body of the coding sequence relative to non-synonymous mutations suggests positive selection on the *trxA* gene. The convergent evolution of Arg74 to Cysteine or Leucine in 3 out of 4 bioreactors further points towards a strong selection pressure on this gene. These signs of positive selection on the *trxA* gene prove that the continuous culture conditions the phage are subjected to are suited to directed evolution. The mutation rate is high enough for mutations to accumulate in the phage genome, and beneficial ones are selected for.

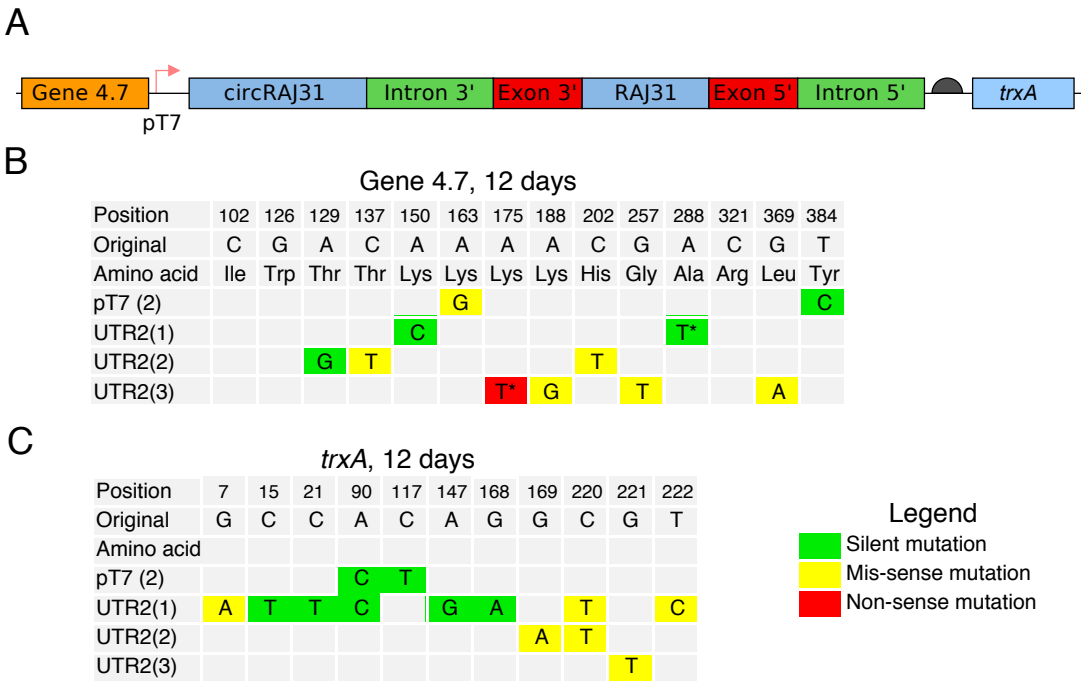


Figure 3.10 – Dominant mutations in T7-circRAJ31 after 12 days of continuous evolution in pT7-*cmk* or UTR2-*cmk* cells. **A**. Fragment of T7 genome sequenced. **B**. Mutations found in coding regions. 4.7 shows no evidence of selection and both silent and non-silent mutations appear. **C**. *trxA* shows evidence of strong selection, with 7 out of 13 mutations being silent and 4 of the 6 non-silent mutations affecting a single amino acid. Asterisks indicate a detectable mixed population (mutant + original) in the final sample.

3.2.7 Mutations in circRAJ31 sequence

Mutations appeared in the circRAJ31 sequence in all 4 bioreactor runs, shown in fig. 3.10C. An average of 7.5 mutations appeared in circRAJ31 when evolved in Nu2_UTR cells, a similar number to when evolved in the negative control cells containing pT7-cmk (8 mutations). The mutations were spread throughout the promoter, ribozyme and RAJ31 riboregulatory region sequences. To determine whether these mutations conferred any advantage to the riboregulator sequence, they were re-synthesised and cloned in the same context as the original circRAJ31 riboregulator, removed from the T7 phage sequence. The evolved circRAJ31 sequences were co-transformed with cis-repressed GFP plasmids used in chapter 2, which are activated by rationally designed linear and circular versions of the RAJ31 riboregulator (see figs. 2.13 and 2.14). The

ability of the evolved riboregulators to activate translation of cis-repressed GFP was then evaluated using the same setup.

A

PIE ribozyme, 12 days

Position	876	879	893	895	918	942	948	971	977	979	989	997	1005		1020	1022	1023	1024	1025		3	7	30	33	58
Original	T	C	G	C	A	T	G	G	A	C	A	G	T		C	G	T	T	T		A	A	T	T	
pT7 (2)					G*	C*									G			C	C				C		
UTR2(1)	C	T	C				A*					A						C					C		
UTR2(2)								A									A	C			G	C		C	
UTR2(3)								A			G		C			A									

pT7, 12 days

Position	3	7
Original	A	G
pT7 (2)		A
UTR2(1)		
UTR2(2)	G	
UTR2(3)		

RAJ31,12 days

Position	1	8	18	22	42
Original	C	G	A	T	A
pT7 (2)			G	C	
UTR2(1)	T	Δ			
UTR2(2)					G
UTR2(3)					

Figure 3.11 – Dominant mutations in the circRAJ31 region of the T7-circRAJ31 phage after 12 days of continuous evolution in pT7-*cmk* or UTR2-*cmk* cells. Asterisks indicate a detectable mixed population (mutant + original) in the final sample.

3.2.8 Characterisation of evolved circRAJ31 sequences

A

Bioreactor host	Evolved circRAJ31
pT7-*cmk*	Evo_circRAJ31_**0**
pT7-UTR2-*cmk*	Evo_circRAJ31_**1**
pT7-UTR2-*cmk*	Evo_circRAJ31_**2**
pT7-UTR2-*cmk*	Evo_circRAJ31_**3**

Resynthesised and cloned downstream of $p_{L\text{-}TetO1}$

Co-transformed with NusI-GFP-LAA

B

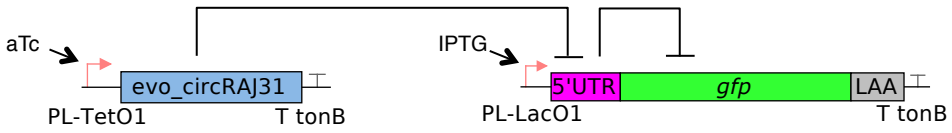


Figure 3.12 – A. circRAJ31 sequences obtained after 12 days of continuous T7 growth in the cellstats were re-synthesised and cloned downstream of

The sequences evolved in samples pT7-*cmk*(2) and pT7-UTR2-*cmk*(1), (2) and (3) were resynthesised in new plasmids, named Evo_circRAJ31_0, Evo_circRAJ31_1, Evo_circRAJ31_2 and Evo_circRAJ31_3 respectively (see fig. 3.12A). They were cloned on the same high-

copy plasmid as circRAJ31, downstream of a $P_{L\text{-}tet}$ promoter, and co-transformed with 3 different versions of cis-repressed GFPs: cisRAJ31, Nusl3 and Nusl6/ This permitted the testing of the circuit with inducers aTc and IPTG controlling the riboregulator and the cis-repressed mRNA (See fig. 3.12B), to determine whether the mutated sequences can control the translation of GFP from a cis-repressed mRNA. The evolved RNAs were found to have lost their activity against all three cis-repressed RNAs tested. When expressed alongside cisRAJ31 (see fig. 3.13A), the evolved RNAs show much lower activation of GFP translation than the unmutated circRAJ31, showing that the mutations only block circRAJ31 activity. This could happen if the mutations inactivated the self-circularisation of the ribozyme region. When co-transformed with Nusl3 or Nusl6 cis-repressed mRNAs, which have a more similar secondary structure to the *cmk* repressing UTRs that were used during the evolution, the evolved circRAJ31 sequences also did not have any activity. These results show that although there was selective pressure for beneficial mutations in the recombinant phage, the mutations that accumulated in the circRAJ31 sequence are most likely random neutral mutations that are unlikely to have been the result of artificial selection for riboregulation of the *cmk* gene.

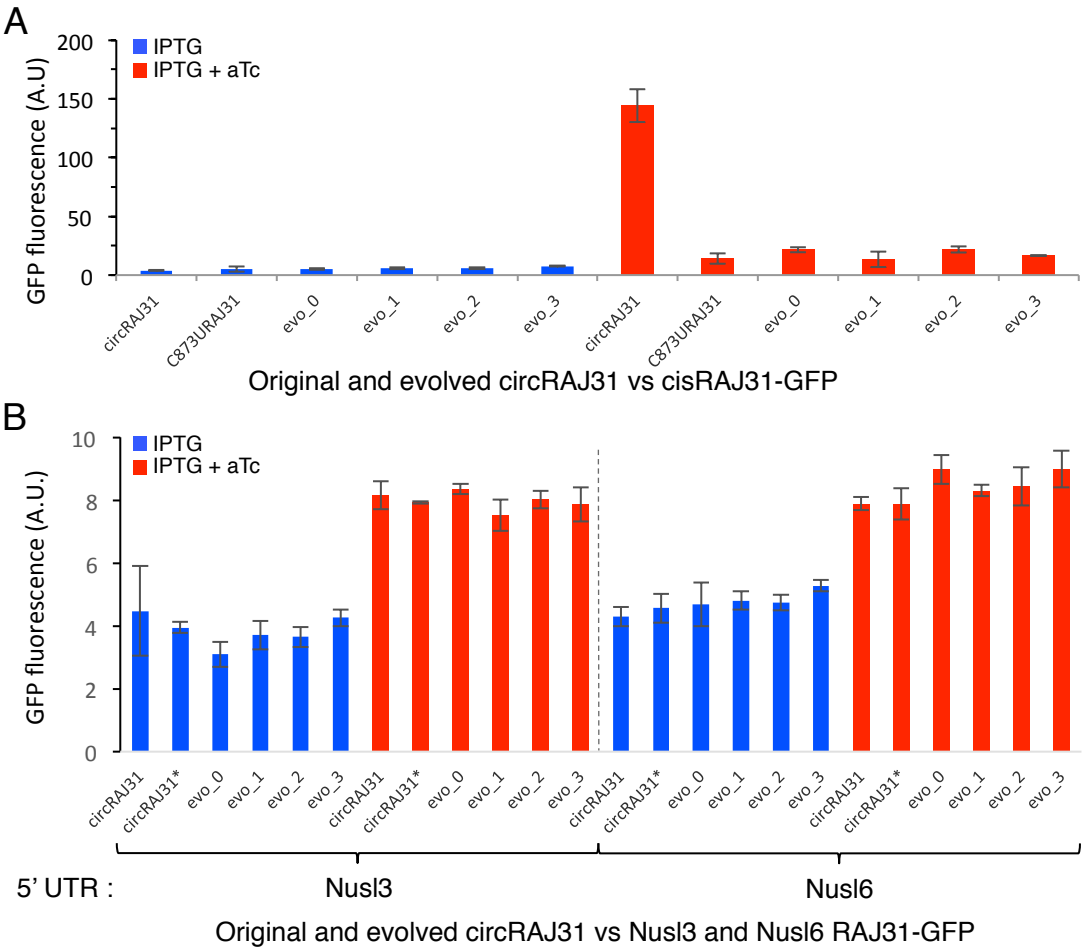


Figure 3.13 – Characterisation of evolved circRAJ31 sequences. **A**. Summary of the sequences obtained from evolution experiments. **B**. Diagram of the circuit used for characterisation. **C**. Fluorescence of the circuits with and without riboregulator expression, using 2 different cis-repressed mRNAs. Evolved riboregulators do not cause an increase in fluorescence when compared to the non-evolved versions.

3.3 Discussion

3.3.1 Recombineering with T7 bacteriophage

We found that recombineering with T7 bacteriophage using *trxA* and *cmk* as selection markers was a viable strategy which allows the production of recombinant T7. Exogeneous genes could be recombined onto the bacteriophage genome by including them within 100 base pair homology arms along with the selection gene, permitting the easy generation of T7 phage carrying foreign DNA. Initially, two selection genes (*trxA* and *cmk*) were used, with the "cargo" in the middle. This was then found to be superfluous and the *trxA* gene was used alone. Homologous recombination was used to insert various genes onto the T7 genome. The longest was *cmk*-λ_{cI}-*trxA*, which is 1839 bp long. As the cassette is inserted in place of T7 *gp5*, which is 2133 bp long with the RBS region included, this modification actually reduced the phage genome size. However, any extra sequence added onto the T7 must fit into the capsid. According to Molineux (2006), the upper limit to this is about 103% of the phage genome size, which means about 1200 bp of genetic material can be added onto the 39937 bp genome, plus whatever is removed in the recombination process. When inserting genes into this locus, the removal of *gp5* provides some extra space, allowing the insertion of 2.9 kbp of foreign DNA, plus the 353 bp required for the *trxA* selection marker. To insert longer fragments of DNA, further elements of the T7 genome would need to be removed. In our case, gene 4.7 (408 bp long) was left on the genome. However, it has been claimed to be non-essential, an assertion that is supported by the accumulation of non-synonymous mutations in our experiments. In fact, all genes from 4.1 to 5.9 (included) are non-essential according to Studier (1981) and Molineux (2006). The first and last ones overlap with essential genes, but removing the non-essential regions would free about 4100 bp, and allow 5 kbp of exogenous DNA sequence to be carried by a recombinant T7 bacteriophage. Further size increases could result from the deletion of other non-essential regions of the T7 genome, which would add up to

about 3500 bp when counting all regions labelled as non-essential by Molineux (2006). Such a phage could carry about 8.5 kbp of foreign sequence.

If carrying out PACE-like directed evolution experiments, this can be compared with the size limit of M13 based phage or phagemids which was used by Esvelt et al. (2011). The size limit for DNA insertion into an M13 phage is about 3 kbp (Brown, 2010), which might be slightly increased due to the deletion of gene 3 (1275 bp). If an M13 phagemid were used for evolution instead, this would free up some space and allow bigger cargo to be evolved. The size limit in this case would be about 10 kbp (Brown, 2010), but with 2.5 kbp taken by the inevitable presence of antibiotic resistance markers, the M13 packaging signal and the plasmid origin of replication, the limit would be about 7.5 kbp. In both cases it seems like slightly bigger sequences could be evolved using T7 as a host rather than M13.

3.3.2 Co-infection and "on the fly" recombineering

The recombination protocol required to engineer T7 is slightly more fastidious than the plasmid-like cloning used for M13. However, we show that this recombination takes place between two phage particles in co-infected *E. coli* cells. We show that this can take place in test tubes, which offers little advantage compared to using a plasmid, but also in continuously cultured phage within a cellstat, in the same conditions used for PACE-like experiments with T7.

The main advantages of using a phagemid carried cassette for recombination lies in the fact that no special procedure is required to make the DNA enter into the cell, as the phagemid capsid contains all the machinery to inject the DNA into the host. In this case, a simple cassette (containing *cmk* and *trxA*) recombines onto the phage genome. This provides an immediate advantage to recombinants, the ability to target a new host population ($\Delta cmk\ \Delta trxA$ *E. coli* cells).

M13 phagemid can also be used to prepare libraries (e.g., Kay et al., 1993). Therefore,

the same recombination method could be used to recombine libraries with an engineered gene during directed evolution. Continuous evolution such as PACE relies on mutations that appear randomly during phage replication. The sequence space accessed is large, estimated to include all possible double mutants in a 1000 bp gene (Esvelt et al., 2011). However, this could be greatly improved if combined with semi-rational engineering techniques that can target libraries to "mutational hotspots" that are known to be important for part function. Such combination of continuous evolution with recombination of libraries could allow big jumps in the sequence space. This would reduce the probability of selecting for a local fitness optimum if more than 1 or 2 mutations are required to improve the part.

Recombineering based methods that allow targeted mutations exist, such as Multiplex Automated Genome Engineering (MAGE, Wang et al., 2009). However MAGE requires the preparation of the cells for electroporation between each round, increasing time and rendering these methods incompatible with continuous evolution techniques.

The ability to bring recombination alongside directed evolution, which itself is a result of the non-disruptive nature of using phagemids to make DNA enter a cell, solves this problem and allows simultaneous continuous evolution and recombination. The advantages of both methods could be combined, continuous evolution of a gene with occasional pulses of diversity generation brought by recombination.

We showed that this type of technology is possible, in this case using cells that are modified to accept co-infection by both M13 and T7. Other phagemids may be also suited, for example phage λ, which can be packaged *in vitro*. M13, λ phage and others can be used to make libraries and could target a continuously evolving gene to increase the search of the sequence space.

The biggest challenge for this to be integrated into continuous evolution experiments is that of the recombination efficiency. Our experiments in non-continuous culture required a "selection" step where non-recombinant phage were eliminated in order to be able to see the recombinant phage. This suggests that recombination is happening

at less than 1 in 10^7, which may be too low to be useful. The recombination efficiency could be increased by adding homology based recombinases into the host cells.

We also found that although co-infection could occur after simultaneous inoculation of cells with both phages in test tubes, no recombinant phage were observed when the M13 phagemid was added into a continuously diluted culture of T7. The larger numbers of T7 in the running cellstat compared to that in test tubes may reduce the chances of M13 from successfully entering the cells and recombining, either because there are less cells or because T7 nucleases are expressed in the cells that the M13 phagemid can infect. In any case, this is not a major issue as infection of the cells in the chemostat permitted recombinant T7 generation. This method could be improved with the use of a delay chamber where the phagemid library is added. This could consist of another bioreactor between the chemostat and the cellstat, or simply of longer tubing between the two reactors. This delay chamber would provide enough time for M13 to enter the cell before the culture is fed to the T7 in the cellstat. As the stock culture of cells in the chemostat would be free of both M13 and T7, successive rounds of infection with different libraries could be achieved.

3.3.3 Picking a selection gene for PACE with T7 bacteriophage

Two potential selection genes for Phage-Assisted Continuous Evolution (PACE) were tested. The first, *gp5*, is the bacteriophage DNA polymerase, whereas the second, *cmk*, plays an important role in nucleotide recycling which allows more phage genome synthesis. *gp5* was found to have little or no effect on phage numbers, despite the fact that DNA polymerase is an essential gene for T7 (Molineux, 2006). The cells containing *gp5* under the control of different promoters and ribosome binding sites were infected with phage containing or not containing the λcI transcription factor DNA. Varying the conditions from which it is expressed, or the presence of the transcription factor had no substantial effect on plaque numbers or appearance. On the contrary, phage infectivity was slightly higher on cells containing the plasmid AP3, which contains

gp5 under the control of a RBS with a very weak predicted translation initiation rate. Although *gp5* expression levels with different RBS and promoters were not directly measured, we can expect that there should be some differences in expression when varying these sequences. The lack of visible differences when changing the promoter and RBS sequences disqualified *gp5* as a useable selection gene for PACE experiments. The initial reasoning behind the choice of *gp5* was that DNA replication would be a key element for production of new infective phage. The lack of decrease in infectivity on cells containing AP3 (predicted weak RBS) compared to AP2 (predicted strong RBS) shows that varying levels of *gp5* did not have the required characteristics, instead leaving to a slight increase in infectivity. The leakiness of the pRM promoter may be sufficient to reach the required levels of *gp5* protein and explain why we do not see any difference.

The testing of *cmk* as a selection gene showed a clear effect on bacteriophage. Plaque morphology was significantly altered, with small plaques on Δcmk cells. Phage killing speeds were affected when *cmk* was placed on accessory plasmids with varying promoter and RBS strengths. This showed that increasing *cmk* levels upon phage entry increased killing speeds, and therefore that *cmk* is a potential good selection gene for T7.

3.3.4 Evolution of *gp4.7* and *trxA*

The random set of mutations with no clear pattern that was found in the various *gp4.7* genes sequenced contrast with the high frequency of synonymous mutations and the convergent mutations that were found in *trxA*. The absence of a significant pattern in the mutations in gene 4.7 suggests a lack of selection pressure on the sequence, with the sequence drifting away in a neutral fashion from the starting sequence. The function of gene 4.7 is not known but it has been reported to be non-essential (Studier, 1981; Molineux, 2006) and that its deletion has no effect upon phage behaviour. Our observation that *gp4.7* accumulates mis-sense mutations in the

bioreactor is consistent with these reports, as there is no apparent selection pressure to maintain the sequence.

On the other hand, an unexpectedly high amount of mutations in *trxA* are synonymous mutations. Furthermore, non-synonymous mutations are clustered in amino acid 74, with two of the four phage populations gaining a Arg74Cys and one gaining a Arg74Leu mutation. Only one in four samples did not finish the 12 day run with a mutation at this site. A high rate of synonymous mutations, as well as the high frequency of this specific mutation at amino acid 74, are both signs of adaptive evolution (Springman et al., 2010). These signs of adaptive evolution in the sequence of the *trxA* gene suggest that the conditions in the bioreactors were favourable for directed evolution experiments. A relatively high number of mutations became fixed and there was convergent evolution of the same mutation in several replicates.

3.3.5 Mutations in circRAJ31 in evolving phage

circRAJ31 accumulated many mutations in the different bioreactors. As it does not contain a protein coding sequence, the comparison of the frequency of synonymous mutations with non-synonymous mutations done for *gp4.7* and *trxA* cannot be repeated. One noted factor was that there were no convergent mutations of circRAJ31 sequences evolved in different bioreactors. Furthermore, the sample evolved in the host containing a pT7-*cmk* accessory plasmid, designed as a negative control where circRAJ31 would not play a role, also had a number of mutations. These mutations were found in all parts of the circRAJ31 sequence: The promoter, the ribozyme introns and exons, and the actual riboregulator sequence that was complementary to the cis-repressed *cmk* UTR targets. After characterisation, the circRAJ31 sequences were found to have all lost their ability to activate cisRAJ31-GFP. The evolution run therefore only led to the accumulation of deleterious mutations in the sequence. This could be because circRAJ31 activation of cis-repressed *cmk* did not lead to a substantial benefit, or because this benefit was lower than the disadvantages conferred by the presence of

a self-splicing sequence within the T7 genome.

3.3.6 PACE with T7 bacteriophage

One of the potential advantages of T7 as a phage for use in PACE-like directed evolution is the fact it is a lytic phage, which avoids clogging and biofilm formation in the cellstat. With M13 phages used by Esvelt et al. (2011), biofilm formation by cells infected with M13 leads to continuous production of phage without any selection pressure, as the cells are no longer continuously diluted, and can lead to cheaters. It was apparent in our bioreactor setting which was easy to run for 6 days without any biofilm growth or clogging. T7 may also have a slightly higher capacity for carrying exogenous cargo, which could be useful when trying to evolve longer sequences. On the other hand, the recombineering with T7 is not as straightforward as M13 cloning, which essentially behaves as a plasmid inside cells. Circuits can be easily cloned onto M13 by standard techniques, treating the phage as a plasmid, before replicating it (Brown, 2010). With T7, the linear and lytic nature of the phage leads to the requirement for engineering by recombination.

Co-infection of T7 with M13 was demonstrated to be possible as a means for 'on the fly' recombineering. This could allow the combination of library based techniques, where an important functional site of a biological part is randomised, and of PACE, where continuous evolution is used to apply selection pressure on the part.
The efficiency of recombination in our setting was however too low to allow this method to work without further work to increase the recombination rate. Recombinant phage was produced at levels that were undetectable without selection and amplification, which corresponds to a ratio lower than 1 in 10^7 recombinant phage to wild type phage. An increase in the recombination rate by using recombination-prone strains could improve this.

Our experiments show that *gp5* is not a suitable gene for use as a selection gene, but that *cmk* seems a better choice. However, it could not be used to evolve the circRAJ31

riboregulator. Other genes might be useable as selection genes with T7 and be better candidates. One possibility is to use gene 17, which codes for the T7 tail fibre that interacts with the target host (Kato et al., 1986). Deletion of gene 17 results in phages that lack tail fibres and cannot infect their target (Springman et al., 2010). This is similar to the strategy used by Esvelt et al. (2011) with the M13 gene 3. Other essential class III genes (genes 6.5+), especially the ones that code for part of the virion capsid (genes 6.7, 8 and 10) and tail (genes 9, 11 and 12) might also be suitable as they are expressed at high levels and are essential for the production of progeny. The disadvantage with these choices is that being class III genes, they are all expressed from pT7 promoters by T7 RNA polymerase. Thus, evolution of parts that interact with the endogenous *E. coli* polymerase might not be possible if the evolved "cargo" is expressed as a class III gene.

This problem highlights a more general disadvantage of using T7 as a host for PACE-style directed evolution experiments, which is that a number of T7 proteins highjack or deactivate the normal *E. coli* machinery. This includes the endogenous RNA polymerase, along with *E. coli* proteins degraded by the T7 proteases. Furthermore, *E. coli* is killed by the process of infection, so evolution of toxic proteins could be advantageous in these conditions, but would be a problem if the evolved parts were used in a different setting.

3.3.7 Conclusion

We found that T7 could be engineered by recombination using *trxA* as a selection marker, allowing the production of phage carrying exogenous genes. The process could also work after co-infection of an engineered bacterial strain with an M13 phagemid carrying the recombination cassette, including in a continuously diluted bioreactor. This could permit the combination of library based selection techniques with continuous evolution as it allows recombination to take place during continuous culture of bacteriophage. Such a system would combine the advantages of semi-

rational mutagenesis approaches such as MAGE, where rationally chosen portions of the genome are randomly mutagenised, and the advantages of continuous evolution carried out with PACE. This might permit the evolution of parts which require a local increase in mutagenesis rate that PACE alone cannot achieve.

Two genes, *gp5* and *cmk* were evaluated as selection genes. While the former did not seem useable, the latter showed desirable characteristics. However, the use of *cmk* to evolve the circRAJ31 riboregulator did not lead to an improved riboregulator. Mutations did accumulate in all genes of the T7 genome, with signs of positive selection in the *trxA* coding sequence. This confirmed that despite the lack of improvement of circRAJ31, the constructed system of continuous culture with T7 bacteriophage was subject to evolutionary selection pressures, and that T7 based PACE experiments should be possible. While the compact genome and lytic life cycle of T7 phage might complicate the directed evolution of single proteins, the evolution of whole phage in continuous culture, for example targeted to different bacterial receptors or strains, could also be a possibility.

3.4 Materials and methods

3.4.1 Cloning

All cloning methods were carried out as described in section 5.3. The plasmid pLitmus-chlor-RFP was built for generic cloning using goldengate and used as a starting point for other pLit-chlor derived plasmids using red/white selection. The *cmk-trxA* cassette was amplified from pRec (Tridgett, 2015) with primers PR1 and PR2 then goldengate ligated into pLitmus-chlor-RFP to generate PR100-*cmk-trxA*. λ*cI* and pT7-circRAJ31 were inserted into the BamHI site of PR100-*cmk-trxA*, which generated PR100-cmk-λ*cI-trxA* and PR100-cmk-λpT7-circRAJ31. *cmk* was removed from the latter as well as from PR100-*cmk-trxA* by cutting both plasmids with BglII and BamHI then self-ligating them, generating PR100-pT7-circRAJ31-*trxA* and PR100-*trxA*.

pSEVA631_RFPcmk_gp5EP was built for generic cloning of parts upstream of *cmk*, for accessory plasmid constructions using red/white selection. Accessory plasmids were then constructed by annealing primers (pT7-*cmk* and SL-*cmk* by annealing pT7_anneal and SL_anneal, respectively) or buying synthetic gblocks from IDT DNA (for UTR1-*cmk* and UTR2-*cmk* and using goldengate to insert them upstream of *cmk*. All products were checked by restriction analysis and sequencing.

3.4.2 Bacterial strains used

Recombination experiments were carried out in MG1655 cells (see section 5.1). Δ*trxA* cells were used for all experiments involving recombinant bacteriophage, unless otherwise specified. They are BWBW25113 *E. coli* and were obtained from the Keio collection (Baba et al., 2006). Their genotype is: *LacI*$^{+}$ $rrnB_{T14}$ Δ$lacZ_{WJ16}$ *hsdR514* Δ$araBAD_{AH33}$ Δ$rhaBAD_{LD78}$ *rph-1*, Δ*trxA* ::KanR), and were transformed with pT7-GP5. Δ*cmk* Δ*trxA* cells are derived from the same strain but with a knockout of *cmk* by *tetR* insertion, and is tetracycline resistant, and were also transformed with pT7-GP5 unless otherwise specified. Co-infection was carried out using a modified MG1655

cells containing an F sure (TetR) plasmid with *pifA::AmpR* deletion, and further deletions to biofilm formation genes (Δ*fim*, Δ*flu*, Δ*matB::P2LUX*). This 'co-infectable' strain is Ampicillin and Tetracycline resistant.

3.4.3 Bacteriophage strains used

The wild type T7 used was identical to NCBI reference sequence NC_001604.1 for the portions sequenced. It was propagated in MG1655 cells and stored at 4°C in 0.2 μm filtered lysate after centrifugation (7000 rcf, 10 minutes). Recombinant phage were produced as described in section 3.4.7, propagated in Δ*trxA* cells and stored in the same conditions as wild type phage. Wild type phage contamination after recombination was checked by plaque assays on MG1655 and by PCR, and removed by serial dilution and repeating plaque assays on Δ*trxA* -G5 cells until wild type phage was undetectable by both methods. T7 phages were sequenced after PCR of the plaques or filtered supernatant with G5F_new and G5R_new, and sequencing the purified PCR products (Sanger sequencing, GATC biotech) using the same primers.

3.4.4 M13 phagemid preparation

M13 phagemids were prepared by co-transforming the appropriate pLitmus derived plasmids (pLitchlor-PR100 or pLitchlor-RFP) with the M13KO7 helper phage. The co-transformed cells were grown overnight then refreshed 50 fold in LB medium and grown at 30°C for 6 hours with chloramphenicol (17.5 μg/mL) and Kanamycin (25 μg/mL). Cells were centrifugated (7000 rpm, 3 minutes) and the phagemids were recovered by filtering the supernatant with a 0.2 μm filter (Sartorius). Phage concentration was measured by serial dilutions of the filtrate, incubation with TG1 *E. coli* cells (10 minutes at 30°C without shaking, 10 minutes with shaking), plating serial dilutions on LB plates with chloramphenicol (17.5 μg/mL), and counting colonies.

3.4.5 Plaque assays

Plaque assays were conducted with the appropriate cells, grown to O.D. 0.2 after refreshing an overnight culture in LB (1:100 dilution). Cells were placed on ice until use if necessary. 100 µL of phage dilutions were added to 300 µL of cells, then plated in 3 mL of soft LB agar (0.75 % agar) onto LB agar plates containing appropriate antibiotics. Plaques were counted after overnight growth at 37°C. To recover phage, the centre of the plaque was picked with a pipette tip, diluted in 0.5 mL LB and filter sterilised.

3.4.6 Co-infection with M13 and 'on the fly' recombineering

Co-infection of Δ*pifA* cells

E. coli F+ Δ*pifA* biofilm deficient cells were grown overnight, then refreshed 1:100 in 20 mL culture and grown to $OD_{\lambda=600}$ 0.2. This was then split into two, with 150 µL of LB or of M13KO7 phagemid added (MOI ≫ 1), incubated at 37°C for 15 min (10 min without shaking, 5 min with shaking at 200 rpm). 100 µL of wild type T7 dilutions (100 PFU per mL), or LB, was added and plaque assays were performed as described in section 3.4.5, on LB plates with or without Kanamycin (50 µg/mL). Plaque assays were incubated overnight and lawns and plaques counted the next day.

Recombination in batch culture

E. coli F+ Δ*pifA* cells were transformed with pT7-G5[Kan]. *E. coli* F+ Δ*pifA* , PR100-*cmk-trxA* were then infected with M13::PR100-*cmk-trxA* phagemid with tranduced colonies selected for on chloramphenicol LB plates (35 µg/mL), providing a positive control for cells tranformed with the plasmid. Both strains were grown overnight and refreshed (1:100 dilution) in LB. At $OD_{\lambda=600}$ 0.5, the strains were placed at 30°C. *E. coli* F+ Δ*pifA* were inoculated with M13::PR100-*cmk-trxA* or M13::RFP (MOI=20), and both strains were inoculated with wild type T7 (MOI=0.01). When the cultures

cleared, they were centrifugated (7500 rcf, 10 minutes), and filter sterilised, yielding the non-enriched lysate. Enriched samples were obtained by inoculating a 100 mL culture of T7 Δ*trxA* -G5 (at $OD_{\lambda=600}$0.2) with 0.75 mL of non-enriched lysate (MOI = 0.5), then centrifugating and filtering the cleared lysate as above. Plaque assays were then performed on T7 Δ*trxA* -G5 *E. coli* cells as described in section 3.4.5.

Recombination in continuous culture

Co-infection and recombination in continuous culture was carried out in a continuously diluted cellstat that was fed by two chemostats in a 50/50 ratio. Tubing and other hardware are described in section 3.4.10 and on the Evoprog constortium website (Polesel et al., 2016). The system was setup to maintain the $OD_{\lambda=600}$ in the chemostats close to 0.6, and allow a dilution rate of 20 minutes in the 2 mL cellstat, as described in figure 3.4. The system was cleaned with 1 M NaOH overnight, rinsed and primed with media before use. Cell cultures of the recombination strain (*E. coli* F+ Δ*pifA*) and the selection strain (*E. coli* Δ*trxA* -G5) were grown overnight, refreshed (1:50 dilution) grown to $OD_{\lambda=600}$=0.4, then placed in the chemostats. Cultures were grown in the bioreactors for 4 hours during which the stability of the $OD_{\lambda=600}$ in both chemostats was verified from waste outputs. Samples were taken just before T7 addition (t=30min), just before PR100-*cmk*-*trxA* addition (t=60min) then every 30 minutes. After sample 5, phage was added in the chemostat containing the recombination strain, and the final sample taken after 16 hours. The cellstat was inoculated by rapidly opening the cap and inoculating it with 10^5 PFU of T7 phage, or 2×10^9 cfu of M13 phage. Samples were taken from the cellstat waste output, centrifugated and filtered. Lysis assays used 10 µL of filtered lysate and 10 mL of recombination or selection strain.

3.4.7 T7 recombineering

All homologous recombination onto the T7 genome was targeted to the *gp5* locus. Homologous recombination plasmids containing the gene to be incorporated onto

T7 was cloned between two 100 bp recombination arms homologous to T7 genes 4.7 and 5.3, on the plasmid pLitmus-chlor (Chloramphenicol resistant, pMB1 origin of replication). The recombination leads to the deletion of base 14336, the first base after the *gp4.7* stop codon, through to base 16468, the last base of the *gp5* stop codon. *cmk* and *trxA* were used as selection markers in initial constructs, with *trxA* used alone in later constructs, as described. Unless otherwise specified, MG1655 cells (*recA*+) *E. coli* cells containing the HR plasmid were grown to O.D.600 0.2 and infected with wild type T7 (M.O.I. 0.1). After ~2 hours, the lysate was filter sterilised and 0.75 mL added into 150 mL of selection strain cells (either $\Delta cmk\ \Delta trxA$ cells or $\Delta trxA$ only cells, transformed with pT7-G5). The resulting lysate was diluted and used for plaque assays on selection strain cells. Single plaques were PCRed used primers G5F_new and G5R_new to check for a correct sized insert and sequenced using G5F_new. Wild type phage contamination, which is carried over from the recombination procedure, was removed by serially diluting the plaque extract in LB and replating on selection strain cells. Absence of wild type contamination was checked by PCR (using primers CIE20 and G5F_new), then confirmed by plaque assay on cells without a G5 plasmid. All recombinant T7 sequences were confirmed by sequencing the gene 5 region PCR fragment.

3.4.8 UTR design

Cis-repressing *cmk* 5'UTRs were designed using a custom NUPACK script (Zadeh et al., 2011). The code was identical to that described in appendix E but using the first 30 bases of the *cmk* sequence 5' instead of the domain gfp variable (see cmk_5'_30bp in appendix C). Stuctures were checked in NUPACK for prediction of correct structure alone and for hybridisation with the RAJ31 sequence in complex.

3.4.9 Kill curve assays

Kill curves were run with refreshed cultures of appropriate cells growth to O.D. 0.15, then kept on ice until ready. 96 well plates were prepared in a cold room, by mixing 180 µL of cells and 20 µL of diluted phage (MOI 10^{-6}). Plates were incubated at 4°C for 5 minutes, 37°C for 2 minutes, then transferred to a TECAN Infinite F-500 plate reader. Measurement cycles consisted of shaking (2 mm, orbital, 3 min), with $OD_{\lambda=600}$ measured every 4 minutes. $OD_{\lambda=600}$ of 3 replicates is averaged with standard deviations shown as error bars. For time to peak O.D., the time at highest $OD_{\lambda=600}$ is shown. The first time points are ignored as the high apparent $OD_{\lambda=600}$ is an artefact due to condensation on the plate lid.

3.4.10 Bioreactor construction

Interconnected bioreactors were constructed using silicone 2-stop tubing (Tygon SI 3350, 1.42mm internal diameter), customised glass bottles (see appendix C), and peristaltic pumps (Watson Marlow 403U). A 1602N Hova-Bator Incubator (G.Q.F. Manufacturing Co, Savanah, GA, USA) was used to maintain a temperature of 37°C. Small parts such as bottle holders were 3D printed with PLA using an Ultimaker 2 3D printer (Ultimaker). 80 mm computer cooling fans with magnets attached were used as magnetic stirrers. The system was maintained under pressure generated by the peristaltic pumps and stirred with a magnet. Detailed construction information can be obtained on the Evoprog.eu consortium website (Polesel et al., 2016).

3.4.11 Continuous culture of T7-circRAJ31 in bioreactors

For continuous culture of T7-circRAJ31, 4 cultures of $\Delta cmk\ \Delta trxA$ cells were grown overnight and refreshed, then grown to O.D. of 0.5 to 0.7. Chemostats were then loaded with 32 mL of culture, and grown in continuous culture for 24 hours before phage addition. Phage was added by loading the cellstats with 20 µL of diluted T7-circRAJ31

phage (M.O.I. = 0.01), left unstirred for 10 minutes before restarting. Bioreactors were run at various dilution rates, which are indicated in figure 3.9. Dilution rate was adjusted during the run on the appropriate day, by adjusting the flow rate on peristaltic pumps. Sampling was taken from the cellstat waste output. 2 mL were filter sterilised and kept at 4°C, and 0.5 mL (unfiltered) used to make a glycerol stock (25%) and stored at -80°C. 1 µL from phage sample each day were PCRed using primers G5F_new and G5R_new, PCR purified and sent for sequencing using G5F_new and G5R_new and PR2, and analysed for mutations.

3.4.12 Calculations of probabilities of mutation in *trxA*

To calculate the probability of a random mutation appearing in the *trxA* gene, a python script was written. The script takes a DNA sequence as an input and goes through it base by base, creating every possible single base mutation and evaluating whether it is silent. The number of possible silent and non-silent mutations are then counted and used to calculate the probability that a random mutation is silent.

3.4.13 Characterisation of mutated riboregulators

The sequences obtained after 12 days of continuous culture of the T7::circRAJ31 bacteriophage in pT7-*cmk* and UTR2-*cmk* samples were resynthesised as minigenes by IDT DNA, and characterised with cisRAJ31, Nusl3 and Nusl6, three cis-repressed GFP genes (see chapter 2). The TECAN fluorescence assays were conducted as described in section 2.4.4, with $OD_{\lambda=600}$ and GFP fluorescence measurements taken during exponential growth and plotted as GFP/$OD_{\lambda=600}$ during exponential phase.

4 RNA sensors

4.1 Introduction

Many RNA sensors exist in nature, and serve to regulate the transcriptome or to defend against foreign sequences that are synonymous with infection (Filipowicz et al., 2008; Jensen and Thomsen, 2012). Some sense generic patterns of RNA, such as dsRNA or A/U rich RNA, while others are specific to a particular sequence. Unsurprisingly, many sensors are protein based, for example Toll-Like Receptors (TLRs) or RNA binding proteins. Others are only RNA based (many bacterial sRNAs) or include an RNA component (e.g. chaperoned sRNAs, microRNAs). Because RNAs naturally base-pair with complementary sequences, their structure offers an evident way of interacting with other sequences specifically. Both RNA only sensors and ribonucleoproteins (RNPs) use this mechanism to interact with their targets. In the case of RNPs, the protein based component often carries out the function of the sensor, after an RNA-based recognition.

Several natural RNA based sensors use a modular architecture, where a generic RNP complex carries out a task after target recognition by an interchangeable, modular sRNA (Meister and Tuschi, 2004; Barrangou et al., 2007). From a synthetic biology point of view, modifying such a system to create sequence-specific sensors would be of great interest. Such a device, if it could be programmed to respond to any RNA and

result in targeted gene expression, could be used in synthetic systems that sense RNA viruses, or the transcriptional state of natural systems. The base pairing properties of RNA is an attractive way of providing programmability, as it can be used alongside computational predictors of folding, as well as RNA design software, to create RNAs with specific structures (Zadeh et al., 2011). However, to use these properties to create programmable RNA sensors, a generalisable architecture of RNA sensors that relies on base pairing must be designed.

One exciting new development in biological research has come from the study of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) (Barrangou et al., 2007; Haurwitz et al., 2010). The CRISPR system is a mechanism of acquired immune response found in bacteria and archaea. After viral invasion, the Cas system integrates parts of the infecting phage DNA into the bacterial genome, between "protospacer" motifs. These then get transcribed from the bacterial genome, and after processing, are used as RNA guides that base pair with their phage targets, followed by interference through cleavage of phage DNA or RNA. In CRISPR type II systems, the interference complex is composed of a two-part sRNA guide which base pairs with its target, and a single protein effector, Cas9, which cleaves the phage DNA. Using the Cas9 protein from *Streptococcus pyogenes*, this system has recently been engineered beyond its natural capacity to cleave DNA targets, allowing the use of CRISPR to control a number of biological reactions including activation and repression of arbitrary genes. Furthermore, CRISPR targets are programmed by its guide RNA sequence (gRNA) through base pairing, which make them an ideal molecule for integration within RNA circuits.

4.1.1 Natural sensors of RNA

Sensors of specific RNAs are needed to accompany the many regulatory roles played by RNA in all domains of biology. The existence of RNA viruses provides a second incentive for the evolution of sensors of RNA, and many pattern specific RNA sensors

are part of the innate immune response in humans. Broadly, both regulatory and immune sensors of RNA, whether they detect endogenous or exogenous RNA sequence, can be split into two groups: pattern sensors, and sequence sensors. Pattern sensors recognise a particular two-dimensional or three-dimensional shape, or a chemical pattern, regardless of the nucleotide sequence. Sequence sensors detect a specific RNA sequence. Many RNA binding domains sense a specific sequence both through the structure it folds into and the bases that compose it, and therefore belong to both groups.

Sensors of specific sequences

Many natural RNA regulators are RNA sequences themselves, and interact with other RNAs through base pairing (see fig. 4.1A). This is, by nature, sequence specific, even if absolute complementarity is not always required and some sensors can tolerate mismatches. In eukaryotes, siRNAs and miRNAs (Meister and Tuschi, 2004; Filipowicz et al., 2008) are two large classes of regulatory sRNA. Their mature forms are generated through different pathways, but once processed, both miRNAs and siRNAs base pair with complementary sequences, then downregulate their targets through repression of translation or degradation, in association with cellular proteins. Many prokaryotic antisense RNAs work in similar ways. In *E. coli*, MicF is expressed as part of the stress response pathway. It downregulates its target, *ompF* mRNA, by binding to its 5' UTR and blocking translation, then encourages its degradation by other factors (Delihas and Forst, 2001). Similar RNAs exist in Gram-positive bacteria and in archaea (Durand et al., 2012; Bernick et al., 2012). These bacterial sRNAs bind to complementary targets through base pairing, but are often found in association with chaperones like Hfq, although some work alone (Repoila and Darfeuille, 2009). Other cellular processes, such as recognition of Shine-Dalgarno sequences by ribosomes, or of codons by transfer RNA (tRNA) anti-codon loops, also function through RNA base pairing, and could therefore be seen as sequence specific RNA sensors in this respect.

Sensors of shapes

Other sensors detect specific patterns of RNA, rather than specific sequences. These are common within innate immune response to RNA viruses, and bind to RNA patterns that are typical of viral infection (see fig. 4.1B). For example, Toll-Like receptor 3 (TLR3) recognises double stranded RNA (dsRNA), whereas TLR7 binds to uridine- and guanosine-rich single stranded RNA (ssRNA) (Jensen and Thomsen, 2012). Upon sensing of their targets, TLRs activate downstream pathways of inflammation and antiviral immune response. Other proteins, such as RIG-I (retinoic acid-inducible gene I product), or MDA5 (melanoma differentiation-associated antigen 5), are also involved in the recognition of viral-associated RNA patterns (Jensen and Thomsen, 2012). Both are activated by dsRNA, with the former also activated by triphosphate ssRNA – an indicator of foreign RNA, as human transcripts are usually capped.

Many RNases are also specific to certain patterns, often as part of endogenous RNA processing pathways. In bacteria, RNase III cleaves dsRNA, RNase H cleaves DNA-RNA duplexes, whereas RNase E has a preference for A and U rich single stranded regions (Deutscher, 2006; Majorek et al., 2014).

Sensors of both sequence and shape

Finally, many proteins recognise specific sequences, but through recognition of both the structure of RNA and of its specific sequence, without any base pairing mechanism (see see fig. 4.1C). Usually, they bind to the specific stem-loop structures of their targets. MS2 coat protein (MCP) and PP7 coat proteins (PCP) are phage proteins that bind to the RNA genomes of the phages that code for them, serving as the basis of the viral capsids as well as acting as translational repressors (Fouts et al., 1997; Lim and Peabody, 2002). MCP interacts with the MS2 stem-loop structure, making contact with the phosphate backbone of the stem, as well as with exposed bases in the loop and in a bulge within the stem (Valegard et al., 1997). Similar proteins exist on other phages of the same family, each with its own target RNA sequence, usually a stem-loop.

These RNA binding proteins can be fused with others to create chimeric proteins that target specific RNA sequences (e.g. Lange et al., 2008; Delebecque et al., 2011).

Other bacterial proteins also interact with their targets through stem loop motifs. The Hfq chaperone, which binds *E. coli* sRNAs, interacts with its targets through conserved motifs (Updegrove et al., 2016). The Cas9 protein of the CRISPR system recognises the Cas9 handle, formed by the crRNA and tracrRNA, with CRISPR family nucleases from different species recognising different sequences (Esvelt et al., 2013).

Some sequence specific RNases are also specific to a sequence and the structure it folds into. One example is Csy4, which is responsible for processing the pre-crRNA of the *E. coli* CRISPR system. It recognises a 16 nucleotide stem loop, making contact with nucleotides and the phosphate backbone before cleaving, and is unable to cleave the similar stem-loop from *Streptococcus thermophilus* (Haurwitz et al., 2010).

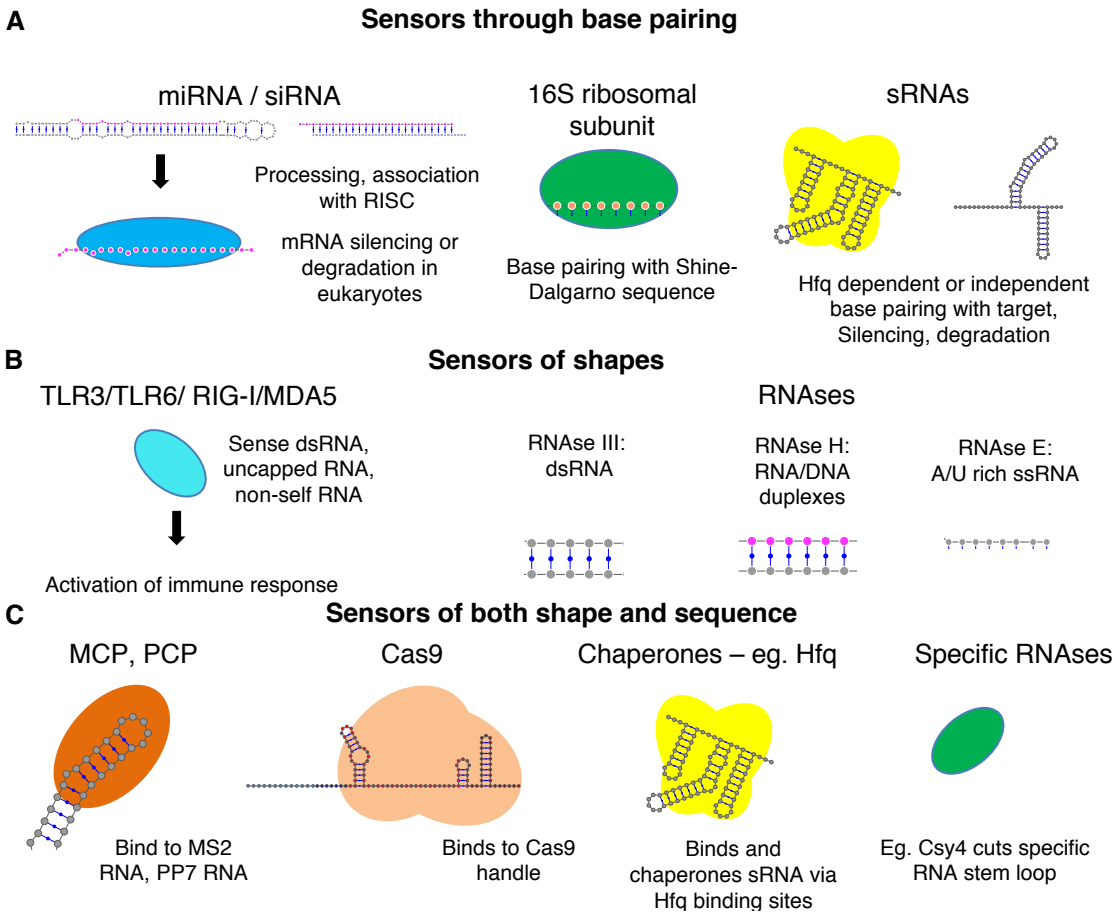


Figure 4.1 – Types of natural RNA sensors. **A**: Natural RNA-based RNA sensors often pair with their targets through base pairing. miRNAs and siRNAs, associate with cellular co-factors such as the RISC complex during processing. They then base pair with their targets, often mRNA, and downregulate it. Bacterial sRNAs also base pair with their targets. The 16S ribosomal subunit interacts with ribosome binding sites through the anti-SD sequence. **B**: Many sensors and RNases are specific to RNA shapes or patters rather than sequence. TLRs and other human innate immune response proteins recognise dsRNA and other types of non-self RNA. Several RNases are specific, like RNase III, RNase H and RNase E. **C**: Some proteins are sequence specific but recognise their target through secondary structure as well as sequence. Examples include MS2 and similar phage coat proteins that bind to RNA loops, Cas9 and the Cas9 handle, or bacterial chaperones. Csy4 is a sequence specific RNase targets a stem-loop.

4.1.2 Synthetic RNA sensors

Synthetic RNA sensors have been engineered, and like natural sensors, they could be divided into pattern-detectors and sequence-detectors. However, most attempts to engineer RNA sensors have focused on the development of sequence specific devices rather than pattern-sensing ones. One exception is the use of RNA binding proteins such as the MS2 coat protein, that sense a fixed sequence, to target modified RNAs. To do this, MS2 binding loops are cloned to the 3' of the RNA of interest, and an MS2-GFP fusion is co-expressed alongside the target mRNA *in vivo*. A similar approach can be taken with PP7, Com, or other comparable RNA binding proteins. This allows the visualisation of RNA expression and localisation in living cells (Bertrand et al., 1998). By repeating this with the PP7 fused with mCherry, different RNAs can be targeted simultaneously (Coulon et al., 2014). Other similar RNA binding proteins can also be used for this purpose (e.g. Lange et al., 2008).

However, the use of these sequence specific sensors require the engineering of the RNA of interest to introduce the MS2 and PP7 binding loops, limiting the use of this strategy to artificial systems where the target sequence can be introduced. Most unmodified sequences do not contain RNA binding protein sites, so this approach is not generalisable. The re-engineering of sensing systems that detect their targets through base pairing is therefore of great interest, as it could theoretically permit the creation of a generic strategy for RNA sensor engineering. Past attempts at creating such generalisable sensors are described below.

Trans-splicing ribozymes

Self-splicing group I intron ribozymes, such as the ones used to create circular riboregulators, can be split into two separate parts by removing the P6 loop, then expressing them from different promoters (Galloway Salvo et al., 1990; Sullenger and Cech, 1994). When transcribed, the two RNA parts still hybridise, and fold into a functional ribozyme. The splicing reaction then ligates the exon sequences, as it does in the wild

type system, resulting in an effective *in vivo* RNA ligation. The ribozyme splicing sites also can be modified, so long as the secondary structure of the P1 and P10 stems, where the splice targets are located, remain the same, and that the 5' splice target is a U base (Köhler et al., 1999). Trans-splicing ribozymes hybridise with their target RNAs at their 5' end – called the internal guide sequence (IGS) – and reform the P1 stem. Complementary changes at the 3' end of the ribozyme allow the formation of the P10 stem. The sequence downstream of the 3' splice site is then ligated onto the target RNA. Trans-splicing ribozymes can therefore be engineered to target a specific RNA sequence, and replace its 3' sequence with an arbitrary one placed on the trans-splicing ribozymes. The specificity and activity of such trans-splicing ribozymes has been improved by the addition of an external guide sequence (EGS), which consists of a 100 nt antisense RNA that binds downstream of the target, bringing the ribozyme and its target closer together (Köhler et al., 1999). Trans-splicers have been used mostly with medical applications in mind, such as to target and repair mRNAs *in vivo* (Sullenger and Cech, 1994; Lan et al., 1998; Nawtaisong et al., 2015). However, they are programmed using the base-pairing rules of RNA, to splice onto their target in a sequence specific way, and are therefore a form of sensor. Furthermore, the cargo which is spliced onto the target could code for various protein outputs, which could in turn activate other circuits.

Artificial ribozyme based RNA sensors

Hammerhead ribozymes (HHR) are self-cleaving ribozymes that were originally discovered in plant viruses and viroids, and are widespread genetic elements found in bacteria and eukaryotes (Hutchins et al., 1986; de la Peña and García-Robles, 2010). Engineering of aptamer or sRNA binding regions into the loop of the HHR stem 3 has allowed the creation of sensor HHRs which are activated or repressed upon small molecule addition (Yen et al., 2004; Klauser and Hartig, 2013). Depending on the placement of the sensor region, the binding of the ligand stabilises or destabilises the ribozyme core, permitting or blocking cleavage. Cleavage releases a cis-repressed

mRNA at the ribozyme 3', or in eukaryotes, encourages RNA degradation. Sensing of the ligand is therefore linked to gene expression. Shen et al. (2015) added an extra layer of complexity, with the "regazyme", a small molecule or sRNA inducible ribozyme which releases another sRNA. This sensor acts as an RNA-based 2 component circuit, with the released sRNA able to activate downstream genes. One advantage of this system is that it is transportable between organisms, as the HHR self-cleaves in prokaryotes and eukaryotes (Klauser and Hartig, 2013). As with trans-splicing ribozymes, the RNA base-pairing rules which govern the programming of such systems are a key element as such devices could in theory be expanded to any organism and to many RNAs.

Toehold switches for mRNA sensing

Toehold switches (Green et al., 2014), described in section 1.2.1, can be targeted to mRNA sequences and act as sensors in *E. coli*. mRNA sensing toehold switches were slightly modified relative to the synthetic trigger-sensing switches, with an increased toehold length (from 12 or 15 nt to ≥ 24 nt), which aimed to decrease the sensitivity of the sensors to mRNA secondary structure. The authors were able to create sensors of a fluorescent protein mRNA (*mCherry*), of antibiotic resistance genes (*cat* and *aadA*), and of a natural *E. coli* sRNA (*rhyB*). The sensors are designed automatically using the NUPACK design suite, and the absence of sequence constraints makes them theoretically adaptable to any target RNA. This adaptability might be constrained by the secondary structure of the target sequence, which could render it inaccessible, or by certain sequences which might disrupt the switch if they interacted with elements such as the RBS. However, when sensing mRNAs, there is usually some flexibility in the choice of the trigger sequence, as different parts of the mRNA can be targeted. The toehold switch design has also been used *in vitro*, using cell free extract, and can be freeze dried on paper to create portable RNA sensors (Pardee et al., 2014). The combination of paper based toehold switches with an isothermal RNA amplification method (NASBA, Nucleic acid sequence–based amplification) and with CRISPR-Cas9

proteins targeting specific strains allowed the creation of femtomolar-sensitive, strain specific sensors of Zikavirus (Pardee et al., 2016). Although toehold switches are probably the most accomplished programmable RNA sensors, through their simple mechanism, high sensitivity and flexible programmability, one important constraint is that they are not portable across organisms. The riboregulator mechanism which they are based on depends on the obstruction of the Shine-Dalgarno sequence and the start codon of mRNAs, which is inherent to bacterial and archaeal mRNAs. In eukaryotes, the riboregulator / toehold switch mechanism for control of translation initiation cannot be directly transposed as initiation of translation is dependent on 5' capping of mRNA (Shatkin, 1976). A second constraint is that their output is limited to activation of protein production encoded by the switch RNA. Although this can indirectly lead to other effects, toehold switches cannot directly regulate host gene expression upon activation.

RNA-targeting CRISPR systems

Since the discovery of the CRISPR-Cas system as an RNA guided programmable deoxyribonuclease, several other CRISPR systems have been discovered. All involve RNA-guided interference, usually directed at phage sequences. Type 1 systems do this through a multi-protein complex, whereas type II use a single protein effector. Within class 2, type II use the Cas9 protein to cleave its DNA target with its two RuvC and HNH nuclease domains. Type V use a single nuclease domain containing protein (e.g., Cpf1, see Zetsche et al., 2015a). Although these systems target DNA, some CRISPR classes can additionally target RNA, such as type III-A and type III-B (Zetsche et al., 2015a; Jiang et al., 2016). The C2c2 protein used by Class 2 type VI systems has been shown to be a single-component CRISPR system that only targets RNA (Abudayyeh et al., 2016). These systems are not true sensors, as they can only downregulate their RNA targets rather than activate genetic pathways upon their detection. However they are RNA-programmable systems that recognise their targets through base pairing.

4.1.3 CRISPR-Cas9

Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a system for adaptive immune response in bacteria (Barrangou et al., 2007). The CRISPR system is composed of a protein 'effector', the Cas9 protein, which contains two nuclease domains that cleave each strand of the DNA after an RNA-programmed targeting. The RNA sequence, called the guide RNA (gRNA), determines what sequence the Cas9 protein binds to, through complementary base pairing with DNA. Differences between gRNA expression in natural settings and engineered synthetic gRNA (sgRNA, although synthetic versions are also referred to as gRNAs), are described in figure 4.3. A generic gRNA is 100 to 105 nt long: 20-25 nt of targeting region, 42 nt of Cas9 handle, and 38 nt *S. pyogenes* terminator (see fig. 4.2). The seed region has been shown to be more important for binding than the 5' end of the gRNA (Jiang et al., 2015). As the gRNA is a small RNA part, its secondary structure and base pairing with DNA are responsible for both the binding of the protein effector and for pairing with the target, and as its sequence and structure are relatively simple, gRNAs are good parts to be incorporated into synthetic RNA circuits.

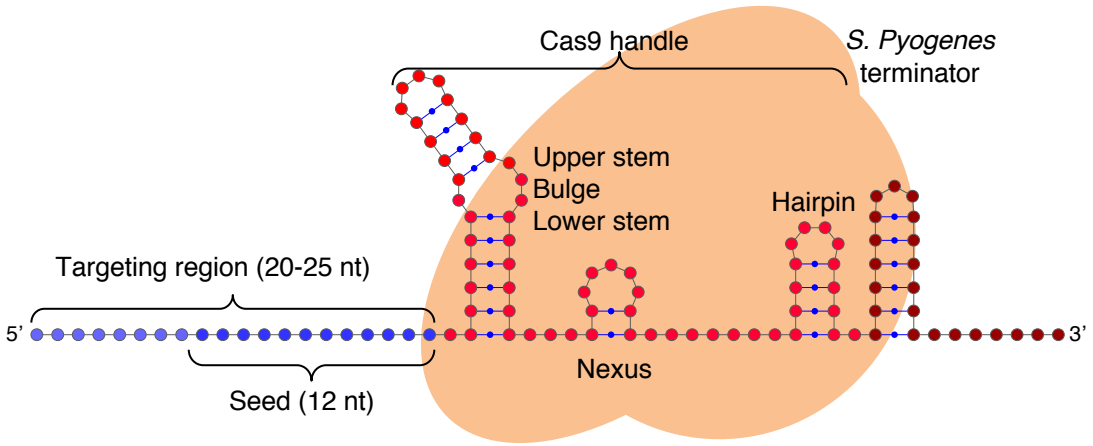


Figure 4.2 – 2D diagram of a sgRNA, with the cas9 protein location on the RNA shown in orange. The blue region is complementary to the CRISPR target, with the 'seed' region shown in a darker blue. The Cas9 handle is shown in red. the *S. pyogenes* terminator used is shown in dark red.

4.1.4 Engineering of Cas9

After RNA-programmed targeting of Cas9 to a specific sequence, the two DNA strands are cleaved by the RuvC and HNH nuclease domains (Jinek et al., 2012). Point mutations deactivating either one of these nuclease domains or both have allowed the creation of nickases (Cong et al., 2013; Ran et al., 2013), and of a catalytically inactive ("dead") dCas9 (Qi et al., 2013). Nickases can be used to engineer eukaroyotic genomes through the homology directed repair system (which is triggered by nicks in the genome) rather than through the error-prone non-homologous end joining repair system (triggered by double-stranded breaks). The catalytically inactive dCas9 was first used by the authors to develop CRISPR interference (CRISPRi). They used dCas and its gRNA to target *E. coli* promoter regions. As the nuclease sites are inactive, the dCas9 remained bound to the DNA, blocking initiation of transcription and effectively acting as an RNA-programmable repressor.

This transformation of dCas9 into an RNA-guided DNA binding protein opened up the path to further innovations, by fusing the dCas9 with other effector domains. As dCas9 had low activity in eukaryotic cells (Qi et al., 2013), the addition of repressor domains permitted the improvement of CRISPRi in eukaryotes. This was achieved with Krüppel-associated box (Cas9-KRAB) and mSin3 interaction domains (Cas9-SID) in humans, or Max-interacting protein (Cas9-Mxi1) in yeast (Gilbert et al., 2013; Konermann et al., 2013). Similarly, the fusion of dCas9 with an Ω subunit of RNAP (in *E. coli*) or with a VP64 domain (in humans) allowed gRNA-targeted activation of transcription, called CRISPR activation (CRISPRa, Bikard et al., 2013; Gilbert et al., 2013).

Other modifications include the creation of a split and photoactivatable Cas9 (Zetsche et al., 2015b; Hemphill et al., 2015), of fluorescent protein tagged Cas9 for imaging of specific loci and the study of chromosome dynamics (Chen et al., 2013), as well as the use of Cas9 proteins from other organisms with differing PAM requirements (Esvelt et al., 2013). These diverse modifications of the Cas9 protein have greatly expanded

the usefulness of CRISPR beyond that of a programmable nuclease, and transformed it into an RNA-guided platform for carrying out many operations in living cells.

4.1.5 Engineering of CRISPR gRNAs

Asides from the aforementioned modifications to the Cas9 protein, the gRNA sequence itself has also been engineered. The first change was the creation of the artificial guide RNA (gRNA), also known as a synthetic gRNA (sgRNA). This was done by fusing the 3' end of the crRNA and 5' end of the tracrRNA through a short loop (Jinek et al., 2012). The two RNAs present in the natural CRISPR system are thus replaced by a single chimeric gRNA, reducing the number of required components to two (Cas9 + gRNA) and removing the requirement for *in vivo* RNA processing (see fig. 4.3). The sgRNA was then modified by extending both stem loops and flipping an A-U base, which seems to increase Cas9 activity in human cells, possibly by stabilising the gRNA (Chen et al., 2013). Further modifications of the gRNA led to the observation that cleavage was reduced if shorter gRNA targeting regions are used. At targeting region lengths between 12 and 16 nt depending on the target, this is due to a lack of cleavage by Cas9, despite the protein binding the DNA correctly (Kiani et al., 2015). This allowed the authors to simultaneously cleave certain sites and activate transcription at others, by using a single Cas9-VPR protein fusion and varying full length and truncated gRNAs. Sites targeted by full length gRNAs get cleaved, whereas those targeted by truncated RNAs are transcriptionally activated by the VPR subunit. The targeting efficiency of truncated gRNAs was also tested using CRISPRi by Qi et al. (2013), who found that repression efficiency dropped to about 15-20% of maximum efficiency using a 12 to 18 nt gRNA, and then was completely abolished with targeting region of 11 nt or less.

Although they did create new gRNA structures, Thyme et al. (2016) showed that the formation of secondary structure in the targeting region reduces or abolishes gRNA activity. The authors tested the ability of several gRNAs to cleave DNA sequences *in vitro* and *in vivo*. gRNAs that contained hairpin sequences in the targeting regions

were inactive, and could be repaired by substituting bases to remove the secondary structure. At least some of these inactive gRNAs were also found to successfully bind to the Cas9 protein *in vitro* and could be used to sequester Cas9 proteins away from active gRNAs, suggesting that the inhibition of the reaction by secondary structure happens at the level of the DNA targeting. Thyme et al. (2016) also found that in cases where the targeting region forms secondary structure with the Cas9 handle of the gRNA, the Cas9 handle can be engineered to remove such interaction. Base pairs of the handle that do not interact with the Cas9 protein (according to the crystal structure) were substituted with another base pair, such that the secondary structure of the Cas9 handle was maintained, but the interaction with the targeting region was removed. This greatly increased gRNA activity, showing both that the gRNA structure affects its activity, and that the Cas9 handle can tolerate modifications. Lee et al. (2016) exploited this property of the gRNA and showed that they could modulate RNA activity by engineering the Cas9 scaffold and introducing secondary structure, this time from an artificial antisense sRNA targeted at the gRNA. The asRNAs had regions complementary either to the gRNA targeting region, or to artificial additions within the Cas9 scaffold, further confirming the flexible nature of the gRNA scaffold, and the possibilities offered by modulation of its secondary structure. Another recent development is that of small-molecule responsive gRNAs, created by incorporating a riboswitch module at the 3' of the gRNA (Liu et al., 2016). This module contains an antisense stem that base pairs with the gRNA targeting region, and an aptamer. In the absence of the small molecule trigger, the binding of the antisense stem with the targeting region inactivates the gRNA. Upon small molecule binding to the aptamer, the antisense stem is displaced, freeing the targeting region. The gRNA is then reactivated and can cleave, repress or activate its target depending on the Cas9 variant used.

The authors built various aptamer-containing gRNAs, allowing the detection of small molecules like tetracycline and theophylline, as well as of cancer-linked proteins such as p53, HSF1 and NF-$\varkappa$B. Since aptamers can be generated *in vitro* by Systematic evolution of ligands by exponential enrichment (SELEX, Tuerk and Gold, 1990; Ellington

and Szostak, 1990), any molecule for which an aptamer can be succesfully created could theoretically be used as an input. As an output, the method offers the same advantages as those of CRISPR, and any DNA sequence could be cleaved, repressed or activated through CRISPR, CRISPRi or CRISPRa.

The Cas9 handle has also been engineered to incorporate MS2 and PP7 protein binding sites in order to permit the engineering of more complex "transcriptional programs" (Zalatan et al., 2015). These stem loops are added to the 3' of the gRNA, sticking out of the Cas9-gRNA complex. A series of fusion proteins were built, containing on one hand an RNA binding protein such as the MS2 coat protein (MCP), PP7 coat protein (PCP) or Com, and on the other hand a transcriptional repressor or activator domain such as KRAB or VP64, respectively. The authors showed that an MS2 gRNA combined with an MCP-VP64 fusion could activate gene expression, and a com-containing gRNA used in conjunction with a Com-KRAB could repress gene expression. This approach encodes the repressing/activating activity in the gRNA rather than in the protein. Using Cas9-VP64 or Cas9-KRAB, the activator or repressor domain would be targeted to any gRNA target present. With this system, one gene can be activated while another is simultaneously repressed, without cross-activation.

Finally, gRNAs have been combined with ribozymes or RNase targeting sites, both to remove upstream and downstream transcribed sequences, and for the multiplexing of multiple gRNAs within a single sequence. Methods have included the use of auto-catalytic ribozymes such as the Hammerhead ribozyme (HHR) or the Hepatitis Delta Ribozyme (HDV), the addition of Csy4 targeting loops to either side of the gRNA, or the addition of flanking tRNAs, which get cleaved by the endogenous RNase P and RNase Z tRNA processing ribozymes (Nissim et al., 2014; Tsai et al., 2014; Xie et al., 2015).

To determine whether sensors could be created using CRISPR-Cas9 based outputs, we designed and tested two strategies for creating sensors. The first strategy is derived from the regazyme system, where a small molecule or a sRNA activates cleavage of an

aptazyme. We hypothesised that gRNAs would be inactive unless released from the aptazyme, and therefore that sensors with a gRNA output could be created if gRNA release was dependent on small-molecule or sRNA induced cleavage of the ribozyme. gRNAs were found to function even without the removal of the 5' elements, which led to a second strategy whereby RNA-binding regulatory elements were placed at the 5' of the targeting region. The hypothesis that artificial 5' elements could block gRNA function by secondary structure was tested, and used to create computationally designed CRISPR gRNA based sensors of RNA sequences.

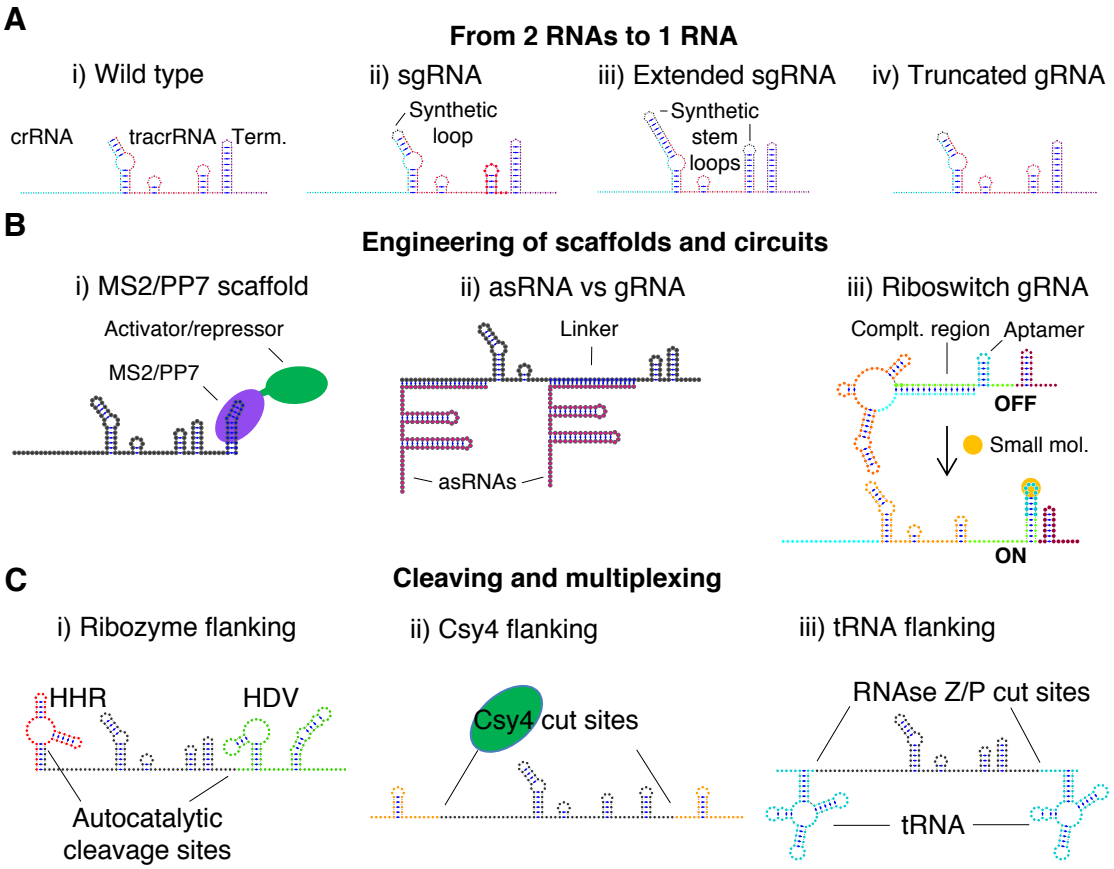


Figure 4.3 – gRNA engineering strategies. **A**: i)Processed wild type gRNA, with a crRNA (cyan) and tracrRNA (Red), and terminator (dark red). ii) Synthetic one-piece gRNA, joined by a loop (black). iii) Extended gRNA with extended stems. Some base pairs can also be replaced. iv) Truncated gRNAs between 12 and 16 bases allow Cas9 binding but not cleavage. **B**: i) Addition of a stem such as MS2 allows the targeting of MCP fused proteins to the gRNA. ii) asRNAs can block gRNAs, by targeting the 5' or an artificial linker. iii) Riboswitch gRNAs can be built by adding small molecule or protein binding aptamers in the gRNA, along with a sequence complementary to the targeting region. The complementary sequence blocks the trigger region, but is displaced by the stabilisation of the stem caused by small molecule addition. **C**. i) gRNA flanked by HHR and HDV ribozymes, which self-cleave and remove surrounding sequences. ii) Csy4 flanked gRNAs are processed by the Csy4 RNase, and can be used for gRNA multiplexing. iii) tRNAs can be used the same purpose, but are processed by endogenous RNase P and RNase Z.

4.2 Results

4.2.1 Design of the 'Crisprzyme'

Shen et al. (2015) previously described the 'regazyme', small molecule or sRNA-activated ribozymes which releases a riboregulator when cut. Regazymes can be responsive to small molecules, such as theoHHAzRAJ11 which is theophylline activated, or sRNA activated, such as breakHHRzRAJ12. Both self-cleave after binding their target. To create sensors of small molecules and RNAs that create an output using the Cas9 protein as an effector, the theoHHAzRAJ11 regazyme was used as a starting point. The aim was to create an aptamer-ribozyme fusion ("aptazyme") that would cleave upon theophylline binding, and release a functional gRNA after cleavage. This provided a simple system to test the compatibility of gRNAs with regazymes, and one that could theoretically be extended to RNA-activated regazymes.

To maintain the structure of the regazyme, the first 25 bases of theoHHAzRAJ11 were kept unchanged. Bases beyond 25 were removed and replaced with the Cas9 handle and *S. pyogenes* terminator (see fig. 4.4A). These changes did not affect the predicted secondary structure of the ribozyme core, which was verified using the NUPACK software. The crisprzyme was thus designed to self-cleave upon theophylline binding, releasing a functional gRNA. An artificial promoter region, composed of the J23119 promoter and the reverse complement of the 25 bases of theoHHAzRAJ11 was cloned upstream of the B0034 RBS and the superfolder GFP gene (see fig. 4.4B). This artificial promoter allowed the core of the crisprzyme to remain unchanged, identical to that of the theoHHAzRAJ11 regazyme, reducing the chance that the modifications would disrupt activity of the regazyme.

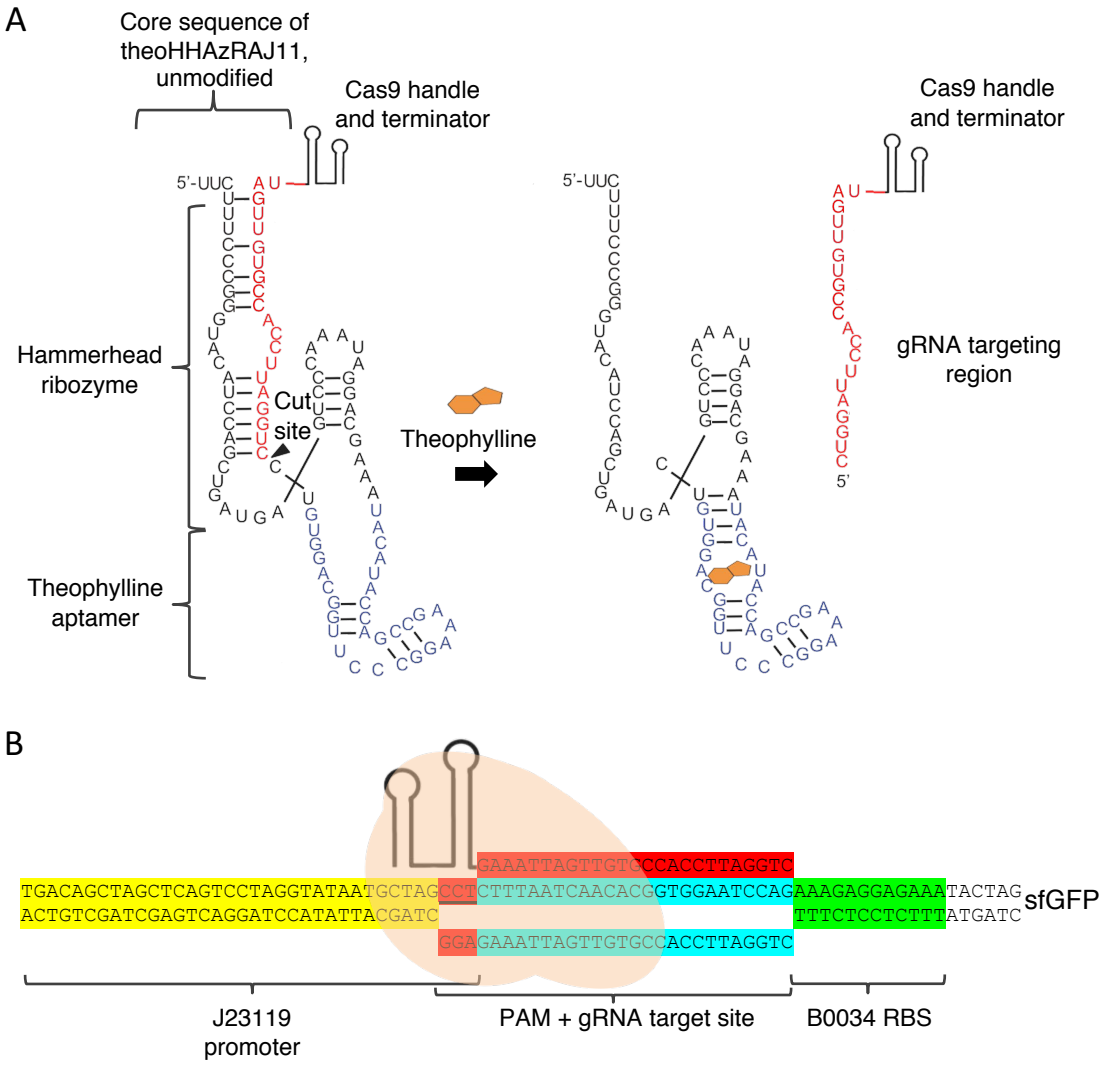


Figure 4.4 – **A**. Design of the CRISPRzyme, based on theoHHAzRAJ11 regazyme (Shen et al., 2015). The core sequence of the regazyme was kept unchanged, whereas the downstream sequence was replaced with the Cas9 handle and *S. pyogenes* terminator. Partially adapted from (Shen et al., 2015). **B**. Promoter reverse engineered to be targeted by the gRNA released by the crisprzyme. The gRNA sequence is shown in red, the Cas9 protein in orange.

A plasmid containing the crisprzyme expressed from a $P_{L\text{-}Lac}$ promoter and an sfGFP gene (Pédelacq et al., 2006) expressed from the synthetic promoter was constructed. This was co-transformed with a second plasmid coding for the dCas9 protein (Qi et al., 2013; Bikard et al., 2013). The circuit was designed to express the crisprzyme RNA upon IPTG induction, and to only cleave upon theophylline addition. A working circuit would repress sfGFP in the presence of both molecules only (see fig. 4.5A). The crisprzyme device was transformed into MG1655-Z1 *E. coli* cells, with or without the dCas9 plasmid. The fluorescence of cells grown in the presence or absence of IPTG (to induce crisprzyme transcription) and theophylline (to induce cleavage) was measured (see fig. 4.5B). A slight reduction of GFP fluorescence (~25 %) was observed when the crisprzyme gRNA was co-expressed with a CRISPRi protein, compared to when the RNA was not expressed. This reduction was not observed in the absence of CRISPRi, suggesting that the gRNA was indeed repressing the transcription of GFP. However, the addition of theophylline to the system when the gRNA was expressed did not change the fluorescence levels. This suggested two possibilities that this assay could not differentiate: Either the ribozyme can cleave without any need for theophylline, or the uncleaved gRNA could repress transcription via CRISPRi, despite the presence of the 79 base pair ribozyme at its 5' end. The reduction in fluorescence levels was also very low, although whether this was due to a weak promoter, an inefficient gRNA target, or to the structure of the RNA cannot be determined by this experiment.

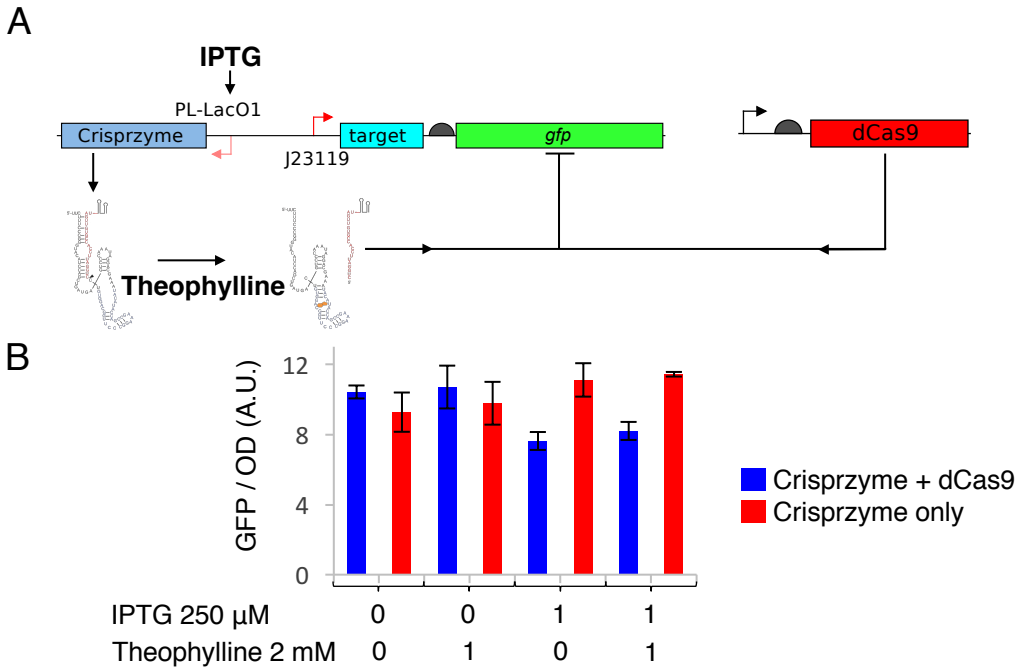


Figure 4.5 – **A**. Circuit used for crisprzyme characterisation. **B**. Fluorescence of *E. coli* cells containing the crisprzyme and reporter system, with or without a CRISPRi expressing plasmid co-transformed.

4.2.2 Uncleaved crisprzymes repress transcription

Two modifications were then made to the crisprzyme. First, the sequence was modified to target a chromosomal pLacI^{q} promoter that controls *lacI* transcription from the *E. coli* chromosome, in order to create an amplifying circuit. Second, an HDV ribozyme was added to the 3' of the crisprzyme to ensure any terminator readthrough sequence was removed (see fig. 4.6A). To establish whether the crisprzymes were cleaving properly and whether they could repress transcription without cleaving, a crisprzyme system was expressed *in vitro* along with a splicing deficient mutant G12A (Ruffner et al., 1990), and cleavage upon theophylline addition was followed over time (See fig. 4.6B). The cleavage products were observed on a gel, which showed that whilst the crisprzyme cleaves upon addition of theophylline, the G12A mutant does not. This confirmed the mutant as an effective non-splicing control. The crisprzyme circuit was then placed under the control of a constitutive promoter, on the same

plasmid as an sfGFP gene under the control of a $P_{L\text{-}Lac}O1$ promoter. The construct was co-transformed with the dCas9 plasmid into *E. coli* MG1655-Z1 cells, which expresses the LacI repressor from a pLacI^q promoter (fig. 4.6C), and the fluorescence of cells expressing this circuit measured (See fig. 4.6D). As the new circuit represses LacI expression, which in turn represses GFP, functional CRISPRi repression results in an increase in GFP. While the pLacI^q was found to have low fluorescence consistent with repression of GFP by LacI, both the crisprzyme and the G12A mutant were found to have high fluorescence, in the absence of theophylline. The G12A mutant, which does not self-cleave, had similar levels of fluorescence as the cleaving Crisprzyme. These results indicated that the uncleaved RNA can still function as a CRISPR gRNA despite the presence of the ribozyme at its 5'.

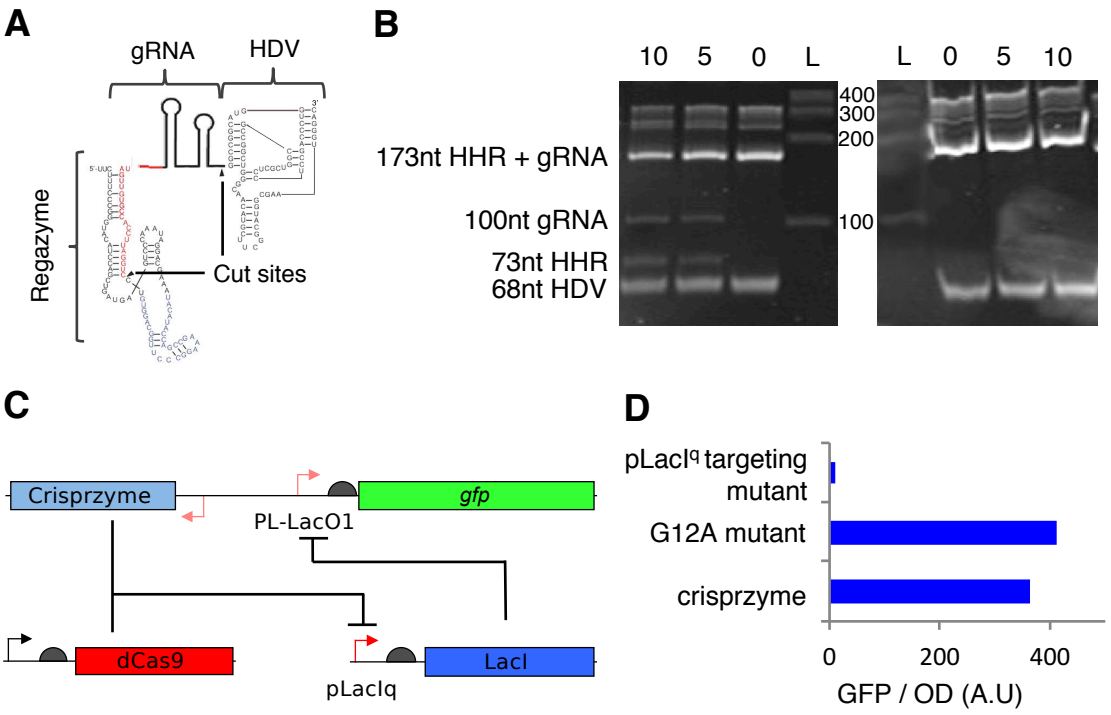


Figure 4.6 – **A**: Modified Crisprzyme with an HDV self-cleaving ribozyme at its 3'. **B**: Cleavage of HHR crisprzyme over time after theophylline addition. *In vitro* transcribed crisprzyme on the left, G12A mutant on the right. L: Riboruler low range ladder. 0, 5, 10: Minutes after theophylline addition. **C**: Modified crisprzyme circuits. The gRNA, in complex with Cas9, represses the pLacI^q promoter on the *E. coli* genome. LacI represses sfGFP. **D**: Fluorescence of *E. coli* cells expressing the pLacI^q crisprzyme, the G12A mutant, or a pLacI^q targeting mutant.

4.2.3 Base pairing in targeting region inhibits gRNA activity

Although the crisprzyme design did not behave as desired, the observation that a long 5' element upstream of the gRNA targeting region does not inhibit gRNA function led to the hypothesis that artificial cis-regulatory elements could be placed at the 5' of an engineered gRNA. For example, if a 5' overhang could be used to inhibit gRNA activity by base pairing, the use of a second regulatory element controls this base pairing by a strand-displacement mechanism (Hochrein et al., 2013; Šulc et al., 2015; Rostain et al., 2015) could be envisioned. To evaluate this idea, the hypothesis that gRNA function could be inhibited by an artificial 5' cis-regulatory structure was tested. A series of gRNAs were constructed with a 5' overhang that was predicted to fold back onto the targeting region (see fig. 4.7). The targeting region of the gRNA is the same as used in section 4.2.2, targeting the pLacI^{q} promoter. The 5' overhang contains a 15 nt loop region, and a clamp region with varying lengths, from 7 to 15 nt.

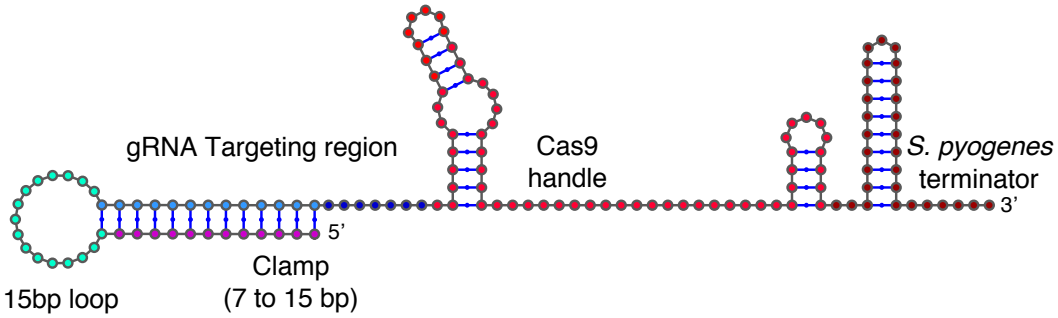


Figure 4.7 – Predicted secondary structure of the self-repressing gRNAs. The 20 nt targeting region (light and dark blue) is repressed by a clamp region (violet), to which it is complementary. A 15 nt loop region (Cyan) separates the two. The clamp was varied from 7 to 15 nt. The Cas9 handle and terminator (red and dark red) are unmodified.

The 15 base pair loop was picked such that it was complementary to an unstructured region of the mKate2 mRNA, and that it did not disrupt the formation of the target structure. To find a sequence that respected these constraints, a custom NUPACK design script was made (Zadeh et al., 2011; Wolfe and Pierce, 2014). The script first folds the mKate2 RNA in a windowed fashion to find unstructured 15 nt regions, then designs a sequence that satisfies the described constraints (see appendix E).

This script was used to design a gRNA with a clamp length of 13 nt. 8 more variants were then built by adding or removing bases complementary to the trigger, representing all 9 gRNAs with clamp lengths from 7 to 15 base pairs. The variants were then re-analysed with NUPACK secondary structure analysis software (Dirks and Pierce, 2003; Dirks et al., 2004, 2007) to ensure correct folding.

The various switch plasmids were co-transformed with a dCas9 containing plasmid into MG1655-Z1 *E. coli* cells. This created a similar circuit to that used in section 4.2.2. An active gRNA-dCas9 complex represses LacI production, and LacI represses GFP production. Thus, active gRNA-dCas9 indirectly induces GFP production (see fig. 4.8A). The fluorescence of cells containing modified gRNAs with different clamp lengths was measured during exponential growth. Negative controls containing only the gRNA or the dCas9 plasmid, and a positive control containing 2 mM IPTG (to inhibit the LacI repressor) were grown alongside. Cells containing shorter clamps had higher fluorescence, showing the dCas9-gRNA complex was repressing the LacI protein. As the clamp got longer, fluorescence decreased, showing a reduction in gRNA-dCas9 repression efficiency (see fig. 4.8B). The highest fluorescence was seen with a 8 nt clamp (around 70 A.U.), higher than the 7 nt (about 40 A.U.). Otherwise, fluorescence was gradually lower as clamp length got longer. The 12 nt clamp had the lowest fluorescence (10 A.U.), and samples with 13-15 nt clamps also had low values (< 13 A.U.). Cells induced with IPTG had high fluorescence (70 A.U.), and cells containing only part of the gRNA-dCas9 complex had very low fluorescence (< 5 A.U.), showing that LacI was repressing GFP production and that its inhibition restores fluorescence. These results indicated that the gRNAs effectiveness at targeting the dCas9 protein could be inhibited by the addition of a cis-repressing regulatory sequence at its 5' end.

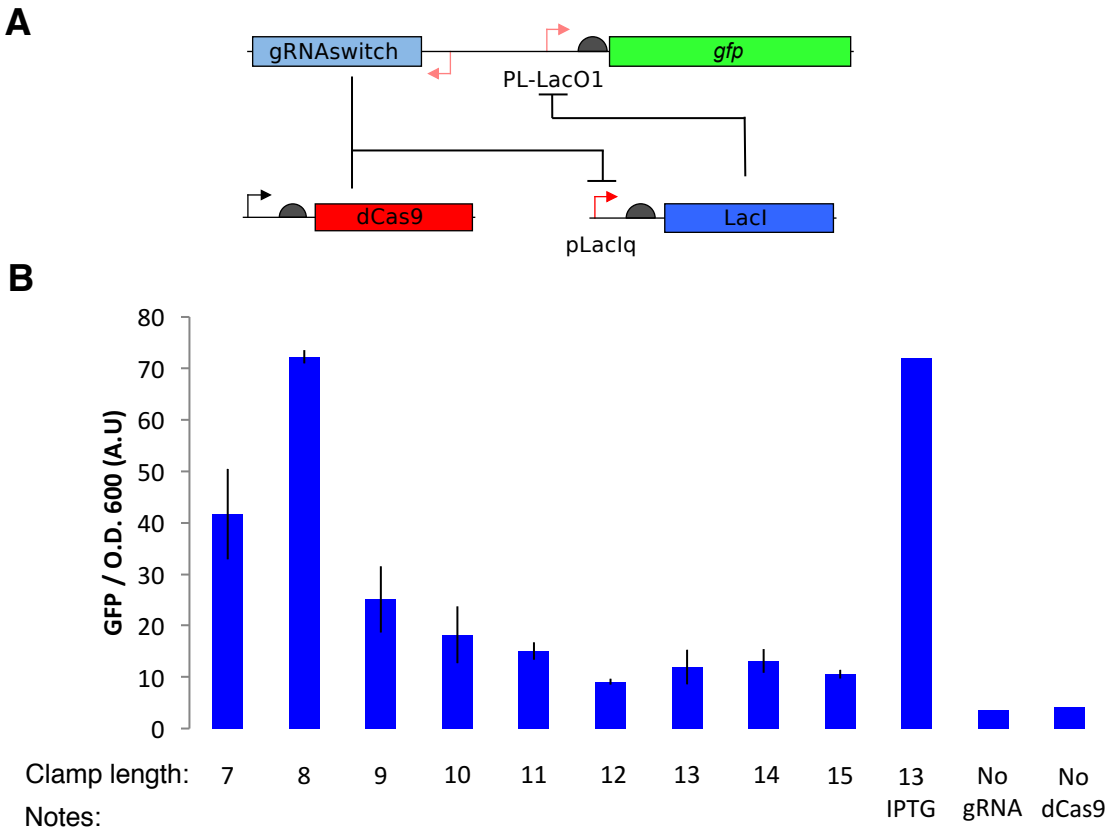


Figure 4.8 – **A**. Circuit used for cis-repressed gRNAswitch characterisation. **B**. GFP fluorescence of MG1655-Z1 cells co-transformed with switches with varying clamp lengths. 'No Cas9' cells contain only the GFP and gRNAswitch plasmid. 'No gRNA' cells contain only the dCas9 plasmid. Sample noted IPTG was grown with 2 mM IPTG. Error bars represent standard deviation between three biological replicates. The fluorescence of samples noted IPTG, No gRNA or No dCas9 are averages of two biological replicates.

The fluorescence of the same circuits was then measured at the single cell level using flow cytometry (see fig. 4.9). The same cells were grown to mid exponential phase and measured for GFP fluorescence. Similarly, the cells were found to have low fluorescence with long clamps, and higher fluorescence with short clamps. Interestingly, cells with medium length clamps had one population with low fluorescence and a tail of cells with higher fluorescence. The declining fluorescence of cells with higher clamp lengths seen in figure 4.8 seems to be due to a mixed population of OFF and partially ON cells rather than to a linear change in GFP expression.

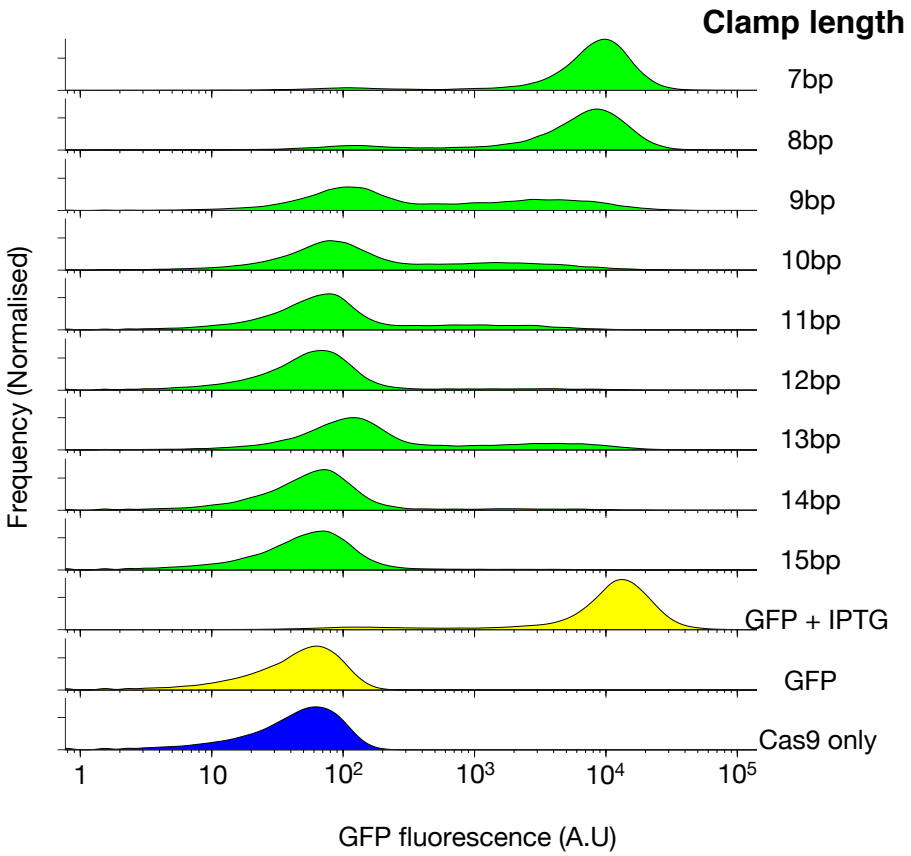


Figure 4.9 – Flow cytometry measurements of GFP fluorescence of MG1655-Z1 cells co-transformed with varying clamp lengths, as described in fig. 4.7. 'GFP' cells (yellow) contain only the GFP and gRNAswitch plasmid, and are grown with or without 2 mM IPTG. 'Cas9 only' cells contain only the dCas9 plasmid.

4.2.4 Design of an RNA activated gRNA

The next step was to determine whether a gRNA with a regulatory self-repressing 5' region (called the "gRNAswitch") could be engineered such that it switches back on after hybridisation with a target RNA (the "trigger" RNA). The gRNAswitch, was designed with a classic gRNA plus a 5' region composed of three elements: The 15 nt loop, a 13 nt clamp, and a 15 nt toehold (see fig. 4.10A and B). The loop and clamp are designed as described above. The 15 nt toehold, like the loop, is complementary to the target RNA. The loop and toehold complement regions on the target RNA are separated by a spacer (fig. 4.10C).

The sequences were designed with a NUPACK script that optimised these structures when the gRNAswitch was alone, but so that the targeting region be unpaired when the gRNAswitch and target are folded together. The co-folded constraints are described in fig. 4.10D. The target RNA is in complex with the loop and toehold region of the gRNAswitch. The clamp region (which is complementary to the targeting region) is not in complex, and forms a bulge with the spacer region of the trigger on the other strand. This binding results in the freeing of the targeting region of the gRNA.

The gRNAswitch 3A was created with this NUPACK script, which accomodates two inputs: An RNA trigger, and a 20 nt CRISPR target. The script then:

1. Creates a trigger and creates complementary toehold and loop regions for the gRNAswitch
2. Creates a clamp of the desired length, complementary to the gRNA targeting region
3. Adds the dCas9 handle and *S. pyogenes* terminator
4. Folds the resulting sequence in NUPACK to analyse their predicted structure

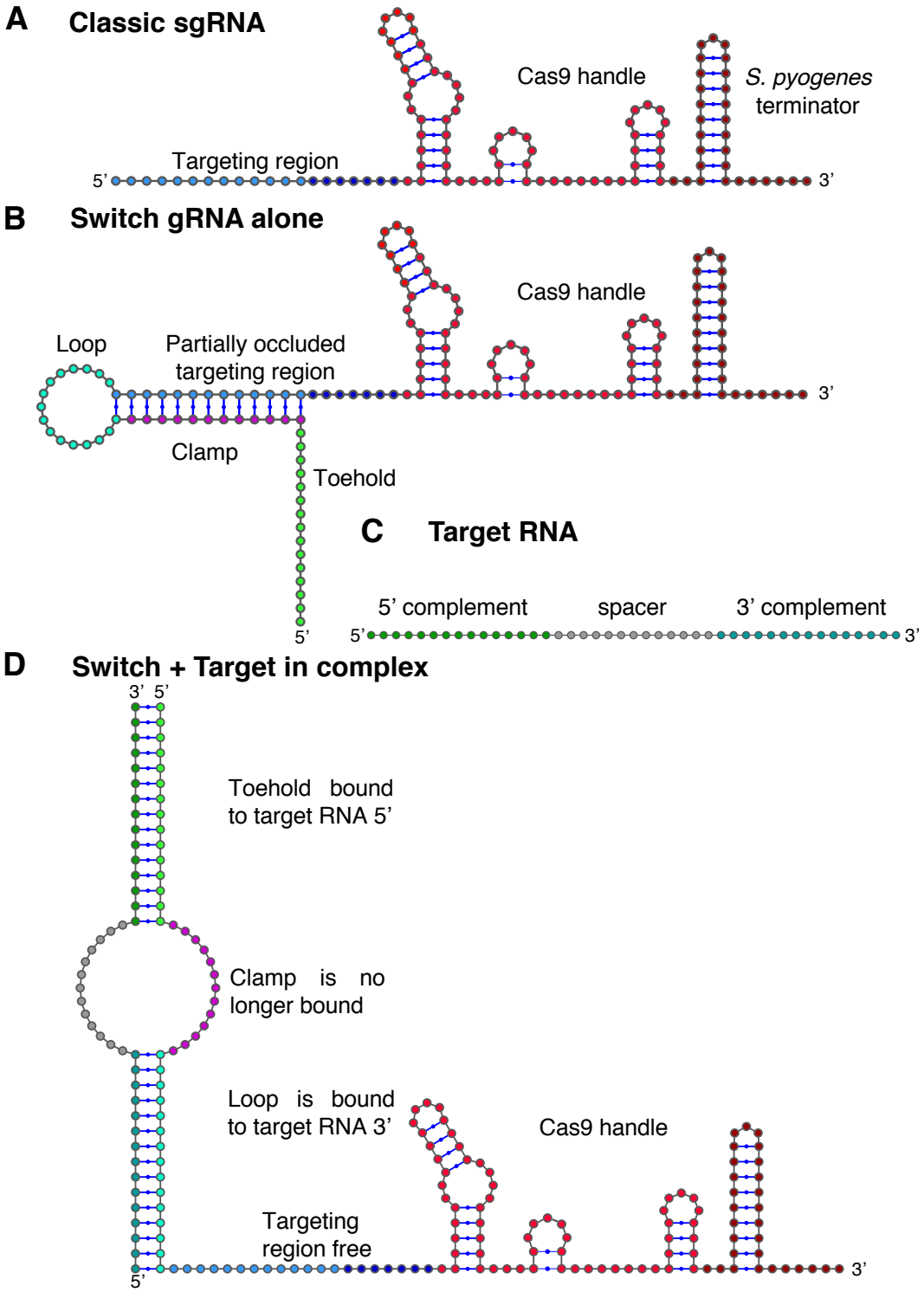


Figure 4.10 – **A**. gRNA secondary structure. **B**. Constraints of gRNAswitch alone. Toehold and loop are single stranded. The clamp region is bound to the targeting region. **C**. The target RNA was chosen to have low secondary structure. **D**. gRNAswitch + target complex design constraints. Toehold and loop regions are bound to the 3' and 5' of the target, respectively. The clamp and targeting region are unpaired.

4.2.5 Trigger 3A can activate switch 3A

The NUPACK script was used to design a gRNAswitch, with a fraction of the mCherry mRNA used as a trigger RNA, and the pLacIq promoter used as a target for CRISPR. The gRNAswitch, called switch 3A, was then tested for its ability to detect a sRNA containing its complement trigger sequence, trigger 3A. A plasmid containing dCas9 and the trigger 3A was constructed (dCas9-trig3A). The switch 3A was co-transformed into *E. coli* MG1655-Z1 cells with either dCas9 (which contains only the dCas9 protein) or dCas9-trig3A (which contains both dCas9 and the trigger). A negative control containing only dCas9, and a positive control containing dCas9 and the trigger3A-switch3A fusion, were also plated. Figure 4.11 shows these cells plated on LB and visualised on a blue light box. Cells containing only the dCas9 and the gRNAswitch were not visibly fluorescent, just as the cells transformed with dCas9 only. Cells containing the gRNAswitch alongside the trigger, however, had high fluorescence, as were cells containing the gRNAswitch-trigger fusion. These results indicate that the trigger can activate the gRNA activity of switch 3A, which is otherwise inactive.

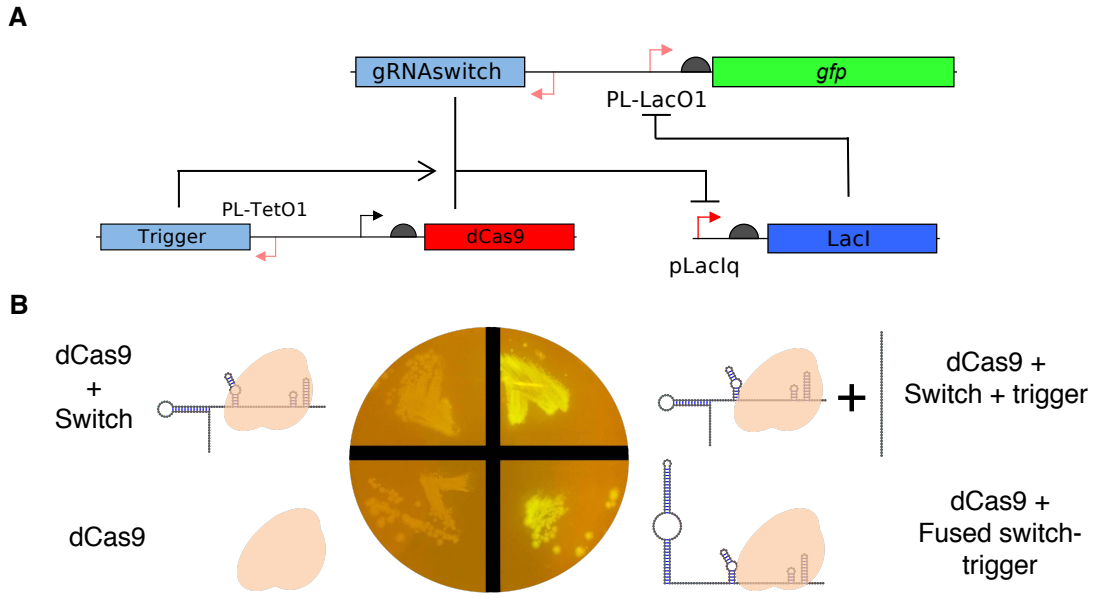


Figure 4.11 – **A.** Circuit used. dCas9 and gRNAswitch are activated by the trigger RNA. When active, they repress lacI, which represses GFP. Functional Cas9 therefore leads to GFP expression. dCas9 has no trigger, nor gRNAswitch/GFP plasmid. dCas9 + Switch has no trigger. dCas9 + switch + trigger has the whole circuit. dCas9 + fused switch-trigger has no trigger RNA, but uses a modified switch where the trigger region is added onto the switch RNA. **B.** Fluorescence of *E. coli* MG1655-Z1 cells transformed with (clockwise from bottom left): dCas9 only, dCas9 and Switch 3A, dCas9, trigger 3A and Switch 3A, dCas9 and fused trigger-switch 3A.

4.2.6 Sensing of full length mRNA

The NUPACK script was then expanded to allow the use of an RNA of any length. In this case the script performs the following, using an RNA of any length and a CRISPR target as inputs.

1. Searches the RNA sequence for unstructured regions to pick trigger regions.
2. Creates a gRNAswitch with toehold and loop regions complementary to trigger.
3. Adds targeting and clamp regions, as well as Cas9 handle and terminator.
4. Simulates folding of the created switches.
5. Picks results that satisfy the constraints best.

This script was used to create two more gRNAswitches targeted to portions of the mKate2 mRNA. Switch 3A-2 and 3A-3 were cloned in place of switch 3A. Two trigger plasmids, dCas9-mKate2 and dCas9-mKate2ΔRBS were made, placing the full length mKate2 mRNA in place of trigger 3A. Switches 3A-2 and 3A-3 were co-transformed with either dCas9, dCas9-mKate2 or dCas9-mKate2ΔRBS in MG1655-Z1 cells. They were grown and their GFP and mKate2 fluorescence was measured in exponential phase (see fig. 4.12). For both gRNAswitches, red fluorescence was high (about 50 A.U.) when mKate2 was present. Deletion of the RBS or removal of the mKate2 gene drastically reduced fluorescence in the red channel (to approx. 5 A.U. or 1.5 A.U., respectively). Switch 3A-2 (fig. 4.12B) had a 2 fold increased fluorescence in the presence of mKate2 mRNA than in its absence (~28 A.U. vs 14 A.U.), regardless of whether an RBS was present upstream of mKate2. With switch 3A-3 (fig. 4.12C), there was no detectable increase in fluorescence in the presence of the mKate2 mRNA. A small increase (from ~9.7 A.U. to 14.3 A.U.) was visible when the mRNA was absent.

Switch 3A-2 was activated by the mKate2 mRNA, increasing GFP fluorescence, although with a small activation fold. Switch 3A-3 however was not functional. The activation of switch 3A by a small RNA and of switch 3A-2 by a full length mRNA showed that the gRNAswitch design can be used as synthetic RNA sensors with a CRISPR-Cas9 protein output.

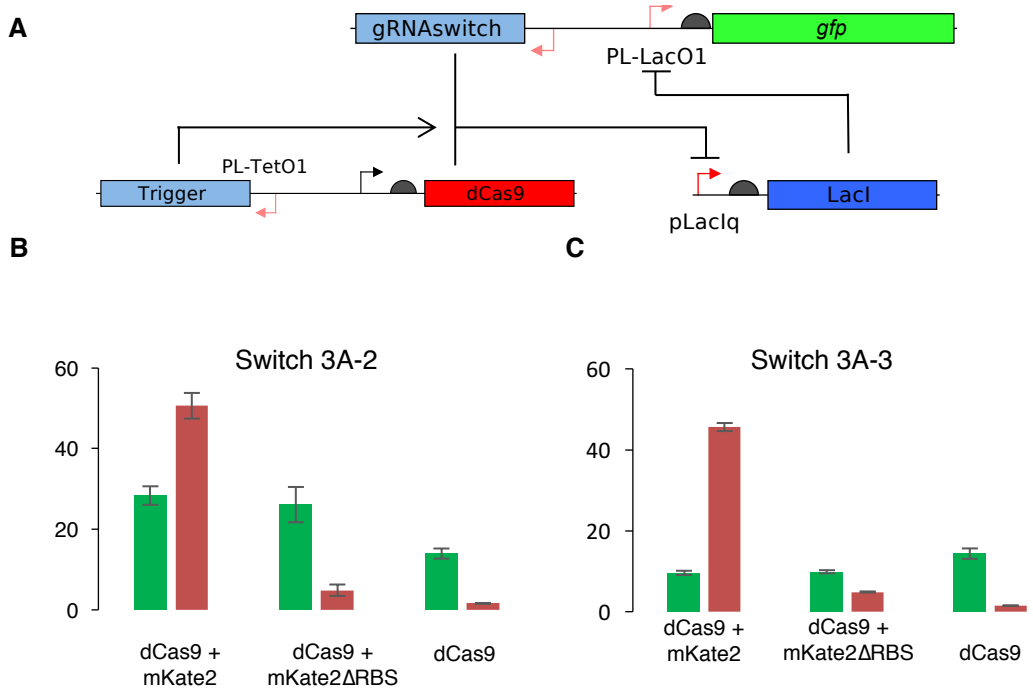


Figure 4.12 – **A.** Circuit used for gRNAswitch characterisation, which should indirectly activate GFP production in presence of the trigger RNA. Here, the trigger is mKate2 mRNA. GFP (green) and mKate2 (red) fluorescence of MG1655-Z1 cells co-transformed with **B**. switch 3A-2 or **C**. switch 3A-3, and dCas9, dCas9-mKate2 or dCas9-mKate2ΔRBS. Error bars represent standard deviation between three biological replicates. The circuit used is identical to that presented in fig. 4.8, but the Cas9 plasmid also contains mKate2 when appropriate.

4.3 Discussion

4.3.1 Crisprzyme activity without induction

The crisprzyme circuit showed a small reduction in gene expression when it was expressed, regardless of the addition of theophylline. It is possible that the crisprzyme might self-cleave even without stabilisation of the theophyline aptamer. Such leakiness has been shown to take place with hammerhead ribozyme (HHR) based theophylline aptazymes (de Silva and Walter, 2009), which both the theoHHAzRAJ11 regazyme and the crisprzyme are. However, the testing of the circuit with the non-cleaving G12A mutant showed that the crisprzyme could repress its target regardless of the removal of the HHR moiety. This defeated the strategy of creating a self-cleaving crisprzyme, as the main assumption it was built on – that the gRNA would not work unless the large RNA sequence at its 5' was removed – was disproved. However, it opened up the new strategy of adding regulatory 5' regions on the gRNA sequences, allowing their integration into strand-displacement based RNA circuits. gRNAs have been modified in their Cas9 handle stem loops (Thyme et al., 2016), between the handle and the *S. pyogenes* terminator (Liu et al., 2016; Lee et al., 2016) and after the terminator in species where it is inactive (Zalatan et al., 2015), but regulatory elements have never been used at the gRNA 5'. These results show that the 5' is also a possible location for adding regulatory RNA structures.

In addition, the use of a ribozyme sequence created a level of unnecessary complexity, with the requirement for correct formation of the ribozyme structure and of cleavage. Finally, the recent development of CRISPR "flow conductors" by Liu et al. (2016) showed that aptamers could be directly incorporated into the gRNAs, solving this problem for small molecules.

4.3.2 gRNA activity is inhibited by 5' cis-repression

The creation of self-repressed gRNAs show that simple 5' motifs complementary to the gRNA targeting region could inhibit a gRNA. The self-inhibition of these gRNAs is coherent with other reports that show that secondary structure formation within gRNAs can interfere with Cas9 mediated cleavage (Thyme et al., 2016), and with the use of trans-acting antisense RNAs that can block CRISPR gRNAs (Lee et al., 2016). Of the gRNAs tested, there was a decrease of CRISPRi activity as clamp length increased, with the exception of the 8 nt clamp which had a higher fluorescence than the 7 nt clamp, close to that of the positive control. This might be due to the 8 nt clamp forming an alternative secondary structure. As these RNAs were constructed by shrinking the clamp region of switch 3A, they were optimised for a 13 nt clamp, which could explain misfolding of shorter clamp lengths. However, the reduction in CRISPRi activity with clamp length shows that artificial secondary structure is an effective way of regulating gRNA activity. Interestingly, there is sharp reduction in activity with clamp lengths of 9 nt or more, which leaves 11 nt or less of the gRNA targeting region accessible. This mirrors the minimum length for CRISPR gRNAs trigger lengths found by Qi et al. (2013), who noted a loss of dCas9 repression when using gRNAs trigger lengths of 11 nt or shorter.

4.3.3 Single cell measurements suggest stochastic changes

The flow cytometry single cell measurements of gRNAs with varying off clamps corroborate the results found in population measurements: fluorescence values are reduced with longer clamp lengths. Single cell measurements did not show differences between the 7 nt clamp and 8 nt clamp, which warrants further exploration. However, a shift towards lower fluorescence was clear as clamp lengths increased. Clamp lengths of 9, 10, 11 and 13 nt all showed a low fluorescence population, as well as a spread of cells with higher fluorescence. This may indicate that the gradual reduction in fluorescence seen in population measurements is due to a mix of GFP producing and non-producing

cells rather than a uniform reduction of fluorescence across the population. The structure of the circuit may explain some of this behaviour. The dCas9 protein and sgRNAswitch are expressed from plasmids with p15a and pSC101-RepAE93K, which have a copy number of 15 and 27, respectively (Hiszczyńska-Sawicka and Kur, 1997; Peterson and Phillips, 2008). On the other hand, the CRISPRi target (LacI) is present in a single copy on the *E. coli* chromosome. In exponential growth in LB, *E. coli* can contain up to 12 replicating chromosomes (Stokke et al., 2012; du Lac et al., 2016), some of which are partially synthesised. Therefore a maximum of 12 CRISPRi target sites can be repressed, or not, by an active Cas9/gRNA complex. The small numbers involved increase the chances of stochastic effects. 12 active gRNAs per cell should theoretically be enough to repress all copies of LacI, and having between 1 and 12 active gRNAs per cell could lead to the intermediate levels seen in some cells.

4.3.4 gRNAswitches can sense RNAs

Cells expressing the GFP circuit and switch 3A were not fluorescent unless trigger 3A was present, showing that gRNAswitches can function as RNA sensors. Switch 3A showed a clear, large increase in fluorescence visible on agar plates. For switch 3A-2, the fluorescence increase was more modest, around 2 fold when measured in a plate reader, but nonetheless this showed that gRNAswitches can sense full length mRNAs as well as small RNAs. Switch 3A-3, designed like switch 3A-2 but targeting a different area of the mKate2 sequence, did not work as an RNA sensor. To determine whether switch 3A-2 and 3A-3 could function as well as switch 3A, they should be tested against sRNA triggers containing their target sequence. However, one possibility for their lower activity is the secondary structure of the mKate2 mRNA. mRNAs are known to have secondary structure *in vivo* (Del Campo et al., 2015), and this could disrupt their correct function. This secondary structure has been an impediment to the development of trans-splicing ribozymes, which also interact with full length mRNAs (Byun et al., 2003; Fiskaa and Birgisdottir, 2010a). To tackle this problem, developers of

trans-splicers have either resorted to a "trans-tagging" assay which uses a randomised trans-splicing library to attempt to identify all possible binding sites (Jones et al., 1996) or have resorted to computational methods (Meluzzi et al., 2012). Here, rather than experimental evidence for lack of secondary structure, the NUPACK design package was used to predict sites with a high probability of being unpaired. However, computational prediction of RNA is known to be imprecise with large sequences, and the possibility that the trigger sites on the mRNA were inaccessible cannot be discarded. The testing of a battery of gRNAswitches against a target mRNA would likely be a better way of finding effective trigger sites.

Aside from secondary structure, the translation of the mRNA by ribosomes could be expected to disrupt trigger-switch interaction. However, for switch 3A-2, the only one to detect mKate2 RNA, deletion of the RBS does not change the GFP output, although red fluorescence greatly decreases.

4.3.5 gRNAswitch characteristics

The gRNAswitch architecture contains 5 non-overlapping modules (Cas9 handle, targeting region, loop, clamp and toehold). It is therefore theoretically adaptable to the detection of any RNA sequence that is long enough to bind, and does not contain strong secondary structure that would block interaction with the gRNAswitch. Sequences that are identical or complementary to the trigger region, or complementary to the Cas9 handle, would also be incompatible with the gRNAswitch architecture. Triggers that interact with the targeting region would not allow the correct unfolding of the clamp, whereas sequences that interact with the Cas9 handle may disrupt Cas9 binding. However, both problems should be rare, and could be solved in some cases by modifications of the switch. If the interaction with the targeting region region is only partial, the gRNA targeting sequence could be modified slightly. gRNAs are known to tolerate a certain number of mismatches, especially in the RNA 5', and changing those bases might remove the unwanted interaction with the trigger, without disrupting

the targeting of the gRNA-Cas9 complex to the DNA. For trigger regions that disrupt the Cas9 handle, then modifications of the Cas9 handle may remove the unwanted interaction. Chen et al. (2013) and Thyme et al. (2016) were able to certain bases of the stem without disrupting Cas9 binding, so long as the secondary structure of the RNA was maintained. Such mutations could abolish unwanted interactions without blocking Cas9 binding. Furthermore, such changes in the Cas9 handle might even be beneficial to the sensing properties of the gRNAswitch. For all tested switches, the Cas9 handle was kept identical to the original single-component sgRNA (Jinek et al., 2012). However, since some base pairs of the stem that do not directly interact with the Cas9 protein can be replaced, these could be designed automatically for optimising the gRNA energetics, rather than being constrained to the original handle sequence.

4.3.6 Potential uses of gRNAswitches

gRNAswitches can detect RNA sequences, using dCas9 as an effector to repress transcription. As the Cas9 handle is not involved in the sensing process, it should be possible to use other variants of the Cas9 protein as effectors upon RNA sensing. DNA cleavage, nickase, CRISPRi and CRISPRa should all be possible outputs after RNA sensing by gRNAswitches. This would enhance their flexibility. These various uses of wild type and engineered Cas9 are known to function in eukaryotic cells, including plants and humans, as well as in *E. coli*. gRNAswitches, however, were only tested in the latter. Their putative mechanism is however generalisable. The base pairing rules that govern RNA secondary structure are the same regardless of the organism, and the mechanism of DNA targeting by Cas9 is also presumed to be identical. If gRNAswitches were functional RNA sensors in yeast and human cells, they could rapidly be used as a research tool, for reporting gene expression or causing cell death upon transcription of a particular RNA.

4.3.7 Improvement of gRNAswitch design

The two functional gRNAswitches designed use 13 nt clamps. However, the clamped gRNAs tested without a toehold showed slightly lower florescence with clamp lengths of 12 or 14 nt. Whether this would also be the case with the toehold-containing gRNAswitches is unknown, and this is in fact likely to vary with the targeting region and with the trigger RNA, which will both modify the structure and the equilibrium free energies of the gRNAswitch. The length of the spacer region, which does not necessarily need to be equal to the clamp length, would also have an effect. To improve the design of the switch, it would be interesting to vary the length of the clamp region in presence of the toehold, as well as to change the lengths of the toehold and loop regions, testing the fluorescence output of the circuit in all cases. The design of gRNAswitches that repress different DNA targets would also be important to prove the flexibility and generalisability of the gRNAswitch design. Flexibility that exists in the Cas9 handle stem could also be exploited, by optimising non-constrained base pairs to create more stable gRNAswitches.

The creation of a library of gRNAswitches with varying clamp, toehold and loop lengths, for several RNA triggers and DNA targets, could allow the identification of the optimal design for these parts. This work has shown that a clamp length of around 12 bases is likely a minimum. As varying all 4 length parameters over large ranges would rapidly lead to untractable numbers, a similar screening should be done for the three remaining parameters of the loop, toehold and spacer lengths to shrink down the space of possible combinations, before a round of screening combining the different possibilities. Although there might be an optimal set of lengths for one gRNAswitch, this is likely to vary when designing switches with different triggers and targets. It would be interesting to study the predicted free energy parameters of the best gRNAswitches, to determine whether there are generalisable design rules pertaining to the length or to the free energy of the different elements of the gRNAswitch, in a similar way to what was done for asRNAs by Hoynes-O'Connor and Moon (2016).

4.3.8 gRNAswitch mechanism

The mechanism by which the gRNAswitch unfolds, although artificial, is not known. The sequences were designed under the assumption that a strand-displacement mechanism could drive such conformational changes and act as a switch. However, it is also possible that active switches are folded into this conformation co-transcriptionally, as shown for riboswitches by Borujeni et al. (2015), rather than re-folded through strand displacement. A second unknown is also whether inactive gRNAswitches are bound to the Cas9 protein, unable to bind their DNA target due to obstruction by the clamp, or whether the clamped gRNA is unable to bind Cas9 altogether. In the context tested, in *E. coli*, these considerations do not affect the behaviour of the switch. However, they might significantly affect the speed with which the sensor could work if used in a different setting. A system which must be switched on co-transcriptionally but can also bind the Cas9 protein in its inactive form would sequester away the available cellular effectors, without acting as a sensor. A trigger RNA could then only be sensed by newly transcribed switch binding to yet unloaded Cas9 protein. However, if gRNAswitches can be reactivated through strand-displacement and bind the Cas9 protein when in their inactive form, they might react faster as they would be pre-loaded and ready to act as an output. Further characterisation is required to better understand the mechanism of gRNAswitches. *In vitro* testing of switch circuits would be particularly useful, as adding the different elements (gRNAswitch, trigger RNA and Cas9 protein) in different orders and observing the output might shed some light on this mechanism.

4.3.9 Conclusion

Two types of CRISPR gRNAs with artificial regulatory elements at their 5' end were tested in *E. coli* using a CRISPRi circuit with a fluorescent protein as an output. The first, called the crisprzyme, contains an aptazyme at its 5' end, which only self-cleaves and only releases the gRNA in presence of theophylline. Although cleavage was conditional to small molecule addition, the unprocessed gRNA could also repress its target even though the 79 nt ribozyme was not removed. This disqualified the crisprzyme design but shows that relatively large regulatory elements can be added to gRNAs in *cis*. This was exploited to design gRNAswitches, cis-repressed gRNAs that are switched off by a clamp region that base pairs with the gRNA targeting region. The gRNAswitch is switched off when using a clamp of 12 nt or more, but is switched back on in presence of its cognate RNA trigger, which is complementary to two other regions of the gRNAswitch, the toehold and the loop.

The gRNAswitch thus functions as a programmable RNA sensor. Although only tested in *E. coli*, the gRNAswitch has no theoretical limitations that would hinder its function in other organisms, including eukaryotes. Furthermore, several paths exist for gRNAswitch improvement, and the many outputs that can be used with CRISPR derived systems should be suited to gRNAswitch outputs, including repression and activation of transcription, DNA cleavage and DNA tagging. gRNAswitches are a new example of the flexibility of the CRISPR-Cas9 system for synthetic biology applications, and shows promise as it combines the programability of RNA circuits with the robust efficiency of a protein based output.

4.4 Materials and Methods

4.4.1 Strains used

All characterisation was carried out in MG1655-Z1 *E. coli* cells. Cloning was performed in TOP10 *E. coli* cells. Strains genotypes are described in detail in section 5.1.

4.4.2 Cloning

Cloning was carried out using Goldengate or Gibson assembly with either PCR products or Gblocks synthesised by IDT DNA. Detailed protocols for assembly methods are described in section 5.3. Primer and part sequences are available in appendix D.

Crisprzyme plasmids

The crisprzyme Gblock was synthesised by IDT DNA and cloned into pSC101-sfGFP-DB by goldengate, resulting in the crisprzyme plasmid (pSC101 origin, KanR). The second crisprzyme plasmid was constructed by PCR and goldengate self-ligation to exchange the promoters.

Cas9 and Cas9 + trigger plasmids

The dCas9 plasmid (ChlorR, p15A origin of replication) was obtained from Bikard et al. (2013). Trigger 3A, mKate2 and mKate2ΔRBS were obtained from Dr. John Duncan. They were constructed by adding the trigger 3A, mKate2 or mKate2ΔRBS onto the dCas9 plasmid by Gibson assembly.

Switch plasmids

Switch 3A was synthesised as a Gblock by IDT DNA and inserted by goldengate into pSC101-sfGFP-DB. It was then used as a template to construct the 9 switch plasmids with varying clamp lengths, as well as switch 3A-2 and 3A3. Switch 3A was amplified

by PCR (Phusion polymerase, NEB) using primer GR21 (as a common reverse primer) and one of GR22 to GR30 (as forward primers) to construct switch3A_OFFclamp_7 to switch3A_OFFclamp_15, respectively; or using primers GR3A-2_fw / rv and GR3A-3_fw / rv to construct Switch 3A-2 and 3A-3, respectively. PCR conditions were according to manufacturer's instructions with 3 % added DMSO. Cycling conditions: 95°, 30sec. 30 cycles: 95°, 15sec; 58°, 20sec; 72°C, 3min. Final extension: 72°C, 10 minutes, hold: 4°. PCR products were then purified with a PCR extraction kit (Thermo Fisher), self-ligated by goldengate and transformed. Correct clones were checked by restriction digestion and sequencing.

4.4.3 Population fluorescence measurements

Population fluorescence measurements of circuits (figures 4.5B, 4.7B, 4.12) were carried out as described in section 2.4.4. Briefly, cells were grown overnight, refreshed 6 hours in M9 medium (1:200 dilution) with inducers as appropriate (aTc: 100ng/μL , theophylline: 2 mM, IPTG: 250 nM), then inoculated into 96 well plates for measurement in a TECAN plate reader. For mKate2 mRNA sensing described in figure 4.12, the protocol was identical save for the GFP gain setting (set to 30), and the additional RFP fluorescence measurement step: Excitation: 580 nm/20 nm, Emission: 635 nm/35 nm, gain: 45). Bar charts plot fluorescence over $OD_{\lambda=600}$ at the closest point to $OD_{\lambda=600}$ 0.3, in mid exponential phase. Error bars represent standard deviations of three biological replicates.

4.4.4 Single cell fluorescence measurements

Single cell fluorescence measurements were performed using a BD LSRFortessa flow cytometer (BD Biosciences). Cells were grown in LB with antibiotics – Kanamycin (17.5 μg/mL), spectinomycin (50 μg/mL), and chloramphenicol (17.5 μg/mL) – and IPTG (1 mM), as appropriate. After overnight growth, cells were refreshed in a 96 well plate (5 μL in 195 μL), grown (4 hours, 37°C, 200 rpm), then refreshed again the

same way. When O.D. reached ~0.3, 10 µL of cells were diluted in 190 µL of pre-chilled filtered (0.2 µMfilter, Sartorius) Phosphate Buffered Saline (PBS), supplemented with 2 mM kanamycin to block further protein synthesis. Cells were then kept at 4°C and fluorescence of single cells was measured. Flow cytometer voltage settings were: Forward Scatter (FSC): 632; Side Scatter (SSC): 237; GFP fluorescence (B485-530/30): 509. FCS data was gated to remove cell debris and fluorescence histograms plotted with Flow Cytometry GUI for Matlab (Steinberg, 2012)).

4.4.5 Streak plate fluorescence

A single colony of *E. coli* containing either dCas9, dCas9 and switch3A, dCas9-trigger3A and switch3A, or dCas9 and switch3A-trigger3A was plated. Cells were grown at 37°C on LB agar with appropriate antibiotics and aTc (100 ng/µL) 16 hours then imaged on a blue light box.

4.4.6 RNA gel

The crisprzyme and the G12A mutant were amplified with primers oJD46f and oJD46r, adding a pT7 promoter upstream of each. The PCR product was used to transcribe RNA (Transcriptaid T7 High-Yield, ThermoFisher), purified in acidic phenol chloroform (pH 4.5, Ambion), ethanol precipitated and quantified with a Nanodrop. 10 µg of each RNA was incubated with 2 mM theophylline (buffer: 10 mM Tris HCl, 2 mM $MgCl_2$). 1 µL of RNA was taken at each timepoint and denatured in 2X RNA loading buffer (65°C, 10 min), then loaded on a polyacrylamide gel (8%) and run at 100 V. The gel was stained with ethidium bromide and imaged with a gel-dock.

4.4.7 Crisprzyme and synthetic promoter design

The crisprzyme sequence was based on the sequence of theoHHAzRAJ11 (Shen et al., 2015). The hammerhead based theophylline aptazyme sequence, as well as the first

25 bases after the cleavage site were kept intact. The bases after that were removed and replaced with a Cas9 handle and *S. pyogenes* terminator (Jinek et al., 2012).

The synthetic promoter was constructed as described, adding the gRNA target site downstream of the J23119 promoter sequence (http://partsregistry.org/Part:BBa_J23119). The resulting UTR sequence was checked for secondary structure around the RBS and ATG and for alternate predicted ribosome binding sites. No strong secondary structure around the RBS or ATG sites was predicted by NUPACK. No alternative ribosome binding sites were predicted by the Salis RBS calculator (Salis et al., 2009). The correct RBS had a predicted TIR of 1998, and the next best initiation site had a TIR of 76 according to the RBS calculator.

4.4.8 Automatic design of gRNAswitches

gRNAswitch sequence 3A was automatically designed using a custom NUPACK script. The script used a clamp length of 13 nt, a toehold and loop length of 15 nt, and a spacer length of 9 nt. The full NUPACK code for design of switch 3A is available in appendix E. For design of switch 3A-2 and 3A-3, the rfp_window.accessible[%] = [50,100] option was enabled, which constrains the choice of the trigger region of the RNA to regions predicted to be at least 50% unstructured.

5 Materials and methods

5.1 Strains used

The strains used in several chapters are described here, whereas more specialised strains are described in the Materials and Methods sections of the relevant chapters. TOP10 and DH5α were used for cloning. TOP10 genotype: F- *mcrA* Δ*(mrr-hsdRMS-mcrBC)* Φ*80lacZ*Δ*M15* Δ*lacX74 nupG recA1 araD139* Δ*(ara-leu)7697 galE15 galK16 rpsL(StrR) endA1* λ-. DH5α genotype: F– *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ*80dlacZ*Δ*M15* Δ*(lacZYA-argF)U169, hsdR17(rK–mK+),* λ–. MG1655 (Genotype: K-12 F– λ– *ilvG– rfb-50 rph-1*) was used for phage recombineering. For inducible system testing, MG1655-Z1 or DH5α-Z1 were used. They are identical to MG1655 and DH5α, respectively, but with a Z1 cassette insertion, which overexpresses LacI, TetR and confers spectinomycin resistance (Lutz and Bujard, 1997).

5.2 Chemicals and reagents

Unless otherwise specified, all chemicals were obtained from Sigma Aldrich (St Louis, USA), all enzymes were obtained from Thermo Fisher Scientific (Waltham, USA), all primers and synthetic DNA – Gblocks and minigenes – were obtained from Integrated DNA technologies (Coralville, USA).

5.3 Cloning

Cloning was performed in DH5α or TOP10 *E. coli*. Plasmids were assembled by standard digestion-ligation, Gibson assembly or Goldengate assembly as described in Materials and Methods sections 2.4, 4.4 and 3.4, with protocols described below. All cloning products were verified by colony PCR or restriction digestion, and by Sanger sequencing (GATC Biotech AG., Constance, Germany).

DNA purification

All DNA miniprep, PCR purification or gel extraction procedures were carried out using either Qiagen or Thermo Fisher purification kits. DNA was eluted in Elution Buffer (EB; 10 mM Tris-Cl, pH 8.5).

Restriction digestion

All restriction digestions were carried out with FastDigest restriction enzymes (Thermo Fisher Scientific) according to manufacturer's instructions. Parts cloned into standard goldengate cloning plasmids and other biobrick compatible plasmids were verified with EcoRI and PstI. Buffers used were FD-green for plasmid verification, standard FD buffer followed by PCR purification when DNA was needed for cloning.

Agarose gels

DNA was visualised in 1% agarose gels in a UV lightbox or for gel extractions, in a blue light box. Gel were stained with Ethidium Bromide or SYBR Safe DNA gel stain (Thermo Fisher).

Standard ligation

Standard ligations were carried out with T4 ligase according to manufacturer's instructions. 3 µL of ligation reaction was electroporated for transformations.

Goldengate cloning

Goldengate cloning (Engler et al., 2008) was carried out using 1µL BsaI (New England Biolabs, Ipswitch, USA), 1 µL T4 Ligase (NEB), 2 µL T4 ligase buffer and 100 fmol of each DNA fragment used (PCR product, Gblock, plasmid or annealed oligos), topped up to 20 µL . Reaction was then cycled in a thermocycler. 30 cycles: 37°C, 2 minutes; 16°C, 5 minutes then 10 minutes, 50°C; 10 minutes, 80°C; 4°, hold. Oligos for goldengate were annealed by diluting them to 0.1 µM in EB, heating them to 96°C for 5 minutes in a heat block, switching it off and letting them cool to room temperature. 3 µL of goldengate reaction were electroporated. Four standard biobrick-compatible goldengate cloning plasmids were built and used for generic goldengate clonong (see section 5.3.1).

Gibson Assembly

1.5X Gibson assembly master mix was prepared with 320 µL 5X ISO buffer, 0.64 µL T5 exonuclease, 20 µL Phusion polymerase, 160 µL Taq ligase (all enzymes from NEB), and 600 µL H_2O and kept at -20°C until use. 5X ISO buffer preparation: 3 mL Tris-HCl pH 7.5 (1M); 150 µL $MgCl_2$ (2M); 240 µL dNTP mix (25mM each dGTP, dCTP, dATP, dTTP) ; 300 µL DTT (1M) ; 1.5 g PEG-8000; 300 µL NAD (100 mM) and 2 mL H_2O. Assembly was carried out with 100 fmol of each fragment (5 µL added in 15 µL mastermix) and incubated at 50°C for 1 hour. 2 µL of reaction were electroporated.

Transformations

Appropriate cells were rendered competent using standard protocols. Briefly, an overnight culture was refreshed and grown to $OD_{\lambda=600}$ 0.4. Cells were spun down and washed twice in 4°C water, and once in 10% glycerol, resuspended in 10% glycerol and kept at -80°C. For transformations, 40 µL of cells were incubated with the DNA mix 15 min on ice, and electroporated in 2 mm wide ice-cold electroporation cuvettes with a BioRad MicroPulse Cell Electroporator on the Ec2 setting (2.50 kV, 1 pulse), diluted in 950 µL SOC medium and incubated 45 min at 37°C before plating.

5.3.1 Standard goldengate cloning plasmids

4 biobrick compatible goldengate cloning plasmids were used for easy construction of parts. pSB1C3-DB, pSC101-DB, pSC101-sfGFP-DB and pLitmus-chlor-RFP. All three plasmids contain an RFP gene (Biobrick part BBa_J04450) between a modified biobrick prefix and suffix containing EcoRI-BsaI and BsaI-PstI, respectively. This permits insertion of one or multiple genes using goldengate cloning and red/white selection. Plasmid maps and goldengte cloning sites are described in appendix A.1.

6 General discussion

6.1 Standardisation of RNA part design

The aim of synthetic biology, and even its definition, has not been clearly agreed on by synthetic biologists themselves (Kitney and Freemont, 2012). Some definitions focus on the functional aspects of building artificial organisms with novel properties, while others focus on the multidisciplinary aspect of a field combining information technology, biotechnology and nanotechnology, or on the role of DNA synthesis (de Vriend, 2006). However, most definitions agree on the fundamental influence of engineering practices like standardisation, and concepts such as decoupling and abstraction hierarchy, on the development of a new approach to genetic modification (Endy, 2005; Müller and Arndt, 2012; Cameron et al., 2014). This was inspired by the incorporation of such practices into manufacturing processes throughout the 19^{th} century, and has led to claims that synthetic biology will be the next industrial revolution (Hellsten and Nerlich, 2011). However, it is still unclear whether such practices can really be applied to biology, and if so, how they can be applied (Kitney and Freemont, 2012). The work presented in this thesis uses three different approaches for engineering RNA parts. The construction of circular riboregulators, described in chapter 2, uses a standardised approach, always using the same PIE sequence for RNA circularisation. The directed evolution presented in chapter 3 is a method

which does not rely on precise understanding of mechanism for action, but requires working parts as a starting point, and a "selection gene" through which evolution can be directed. Finally, the strategy used for the engineering of gRNAswitches in chapter 4 is based upon a rationally designed mechanism of action, but with the use of computational models to predict RNA folding and choose some part elements. These three approaches are complementary, and have been used to varying degrees by researchers in the field of synthetic biology.

6.1.1 'Plug and play' RNA parts

The goal to achieve standardisation of 'plug and play' parts has been pursued by some synthetic biologists (Müller and Arndt, 2012), most notably by the Registry of Standard Biological parts (http://parts.igem.org) who have encouraged the spread of the RFC 10 standard for cloning compatibility for biological parts. In the Registry, each part is presented, at least in theory, as an independent, characterised unit. Published literature also includes standardised plasmid collections (Shetty et al., 2008), as well as standardised and characterised promoter and RBS libraries (Kelly et al., 2009; Anderson et al., 2010b). This 'plug and play' approach is not straightforward to implement, as many sequences that are sometimes defined as 'parts' are known to in fact behave differently depending on the context they are used in. This is the case for promoters and RBS sequences (Ellinger et al., 1994), and the pertinence of the standardised, modular approach has been questioned (Kittleson et al., 2012; Kosuri et al., 2013). In this work, the PIE ribozyme method used for generating circular RNA was used as is, without further changes to adapt it to its context. Despite this, it maintained its ability to circularise different riboregulators, suggesting PIE circularisation is a robust method for circRNA generation. This robustness is also backed up by its uses in other studies in *S. cerevisiae* and *E. coli* (Ford and Ares, 1994), to circularise mRNAs or aptamers (Perriman and Ares, 1998; Umekage and Kikuchi, 2009). It is also easy to implement, and so far seems to be a suitable 'plug and play' circular RNA generator.

Although this view of biological parts as isolated units should probably be taken with some skepticism, the other uses of ribozymes throughout the literature indicate that they might be relatively well suited to this 'plug and play' approach. Along with the circRNA generators, the crisprzymes and gRNAswitches presented in chapter 4 also contained ribozymes. The former were derived from a Hammerhead ribozyme with a modified stem, and both systems were transcribed with an HDV ribozyme at their 3' end to ensure that the correct length gRNAs were produced *in vivo*. The ability of these ribozymes to retain their activity in these different contexts, as well as in various other applications Fiskaa and Birgisdottir (2010b); Serganov and Patel (2007); Gao and Zhao (2014), might be due to the fact that ribozymes are stabilised by tertiary interactions (Batey et al., 1999; Woodson, 2015). This contrasts with parts such as RBSs which have low structure, and might therefore be more easily disrupted. The effect of context on the ribozymes should however be measured, as this might reveal context dependent reduction of the splicing efficiency for example, when the techniques used here are not quantitative. Some factors such as those discussed in section 2.3.2 could upset view of PIE ribozymes as 'plug and play' RNA parts, which will need to be confirmed by further use of PIE ribozymes.

6.1.2 Prediction-based design of RNA parts

The 'plug and play' approach has been contrasted with prediction-based design and library screening/selection methods (Kosuri et al., 2013). In this work, we used prediction-based design for the creation of new cr-mRNAs, and for making gRNAswitch RNA sensors. The reverse engineering of the Nusl3 and Nusl6 5' UTRs presented in chapter 2, that are activated by the linear riboregulator taRAJ31min, validates the design goals developed by Green et al. (2014) for designing new cr-mRNAs. The gRNAswitch RNA sensors presented in chapter 4 are a new example of this approach, where design goals for a modified CRISPR gRNA are fixed, and software is used to pick a fragment of a target mRNA which satisfies these constraints. Like toehold

switches, the gRNAswitch design goals are modular, with non-overlapping regions with determined functions. The clamp and targeting regions are dependent on the Cas9 target, whereas the loop and toehold regions are picked by the NUPACK software such that they are both part of a target RNA sequence and fit the desired constraints. The design goals of gRNAswitches also have some similarities with those of toehold switches. Most notably, both use long single-stranded toehold regions (15 nt). This contrasts with the 6 nucleotides used by Rodrigo et al. (2012), and by many natural systems containing YUNR motifs. One difference between the architecture of gRNAswitches compared to toehold switches or other previously designed RNA parts is that gRNAswitches contain two separate target binding regions. One is situated in the 5' toehold and the other in the loop. These are separated by the 'clamp' region which is bound to the targeting region in the OFF state, and trapped in a bulge in the ON state. This separation of the hybridisation domains of RNA parts into two or more distinct regions might be useful in offering more flexibility when designing other RNA parts.

The limits of the prediction-based design used here are highlighted by the low activation of the Switch 3A-2 when sensing a full length mRNA, and by the failure or lower activity of circular riboregulators. For low activation of switch 3A-2, low mRNA accessibility might be responsible, although only three switches were tested, which is too limited a sample to make definitive conclusions. As mRNAs are longer than the sRNAs otherwise considered, their structure is less well predicted by secondary structure prediction software and picking target sequences based on these software is less reliable. The inability of the current algorithms to predict interactions between circular RNA and their targets is another downside of prediction-based design of RNA parts, which cannot efficiently predict pseudoknotted structures.

Despite these limitations, the field of RNA engineering is one that lends itself well to the application of these concepts of prediction-based design. This is permitted by computational models of RNA folding which, even though they are simplified, allow

the design of RNA parts with a desired structure. Combined with rationally designed RNA structure objectives, such as those of previously described riboregulators and toehold switches (Rodrigo et al., 2012; Green et al., 2014), or those of the gRNAswitches described in this work, functional RNA parts can be designed computationally. So far, this approach has only been used for relatively short RNAs, generally not exceeding 100 nt in length, and which do not contain pseudoknots. As RNA-structure prediction models improve, they might allow improved design of pseudoknot-containing circuits such as those involving circular RNAs, or allow reliable prediction of mRNA unstructured regions. Secondary structure prediction software can also be used in combination with experimental methods for RNA structure determination such as SHAPE-Seq (Watters et al., 2016). These methods can be used to determine the real structure of working RNA parts, which can in turn help to improve the design principles used to construct RNA parts (Takahashi et al., 2016). Eventually, biophysical models which predict the tertiary structure of both proteins and RNA may replace secondary structure models. In the meantime, the combination of secondary structure modelling with design rules will remain a useful method for creating new RNA parts.

6.1.3 Evolution of RNA parts

To improve upon parts whose structure is not predictable computationally, we tested a system for directed evolution using continuous culture of T7 phages. Although the phage system allowed the accumulation of mutations in the circRNA, and there was evidence of positive selection on the *trxA* gene inserted onto the phage, there was no sign of positive selection for functional circRAJ31 production. In fact, all the mutations deactivated circRAJ31 functionality. This highlights the difficulty in creating good selection systems in a continuous evolution setup, and the possibility for the organism to evolve towards different objectives than the system was designed for.

Overall, this failure to evolve circRAJ31 in a continuous T7 evolution setup highlights issues with this particular directed evolution setup rather than inform us about the

incorporation of directed evolution methods in RNA part design. The basic idea that creation of computationally designed parts could serve as a starting point for directed evolution could still hold, although a system where testing of the part in the context of the evolution system is more straightforward should be used. These lessons apply to all directed evolution systems, regardless of whether using RNA or proteins.

6.2 Combination of RNA circuits with protein based outputs

The development of gRNAswitches presented in chapter 4 shows that RNA circuits can be combined with a protein based output platform. These results join the list of previously published systems where dCas9 is used with engineered gRNAs. Zalatan et al. (2015) constructed gRNAs containing an MS2 stem loop or similar elements, which allows the recruitment of activators or repressors, whereas Liu et al. (2016) developed gRNA based small molecule sensors. This approach has the advantage of harnessing the programmability and predictability of small RNA parts, with the robustness of the protein outputs. Other RNA-protein complexes have also been used, for example by Sakai et al. (2014) who added Hfq binding sites on riboregulators. However, Hfq only chaperoned the sRNAs, helping the riboregulator activity, but not carrying out the programmed task. Delebecque et al. (2011) also made use of RNA-protein complexes, using RNA as a three-dimensional scaffold for the spatial organisation of hydrogen producing enzymes. In this case, proteins carried out the programmed task (hydrogen production), but RNA was limited to its role as a scaffold and did not interact with other RNAs, or switch states. In the gRNAswitch system, we use RNA for the tasks it is most suited for: interacting with the trigger RNA and with the DNA target in a programmable fashion, and switching states. The output is then carried out by the Cas9 protein, which can also be chosen to act as a nuclease, repressor or activator, depending on the variant chosen. This combination of RNA based computation and protein based actuation might be a good option which combines the programmability

of RNA with the strong activity of proteins. To make these more reliable, it might be useful to measure the parameters of RNA folding in presence of the Cas9 protein. Association rates of CRISPR gRNA with DNA have been measured and modelled (Farasat and Salis, 2016), and if combined with refined models of the folding of gRNA 5' regions, ease the development of CRISPR based RNA sensors.

6.3 Uses of RNA circuits

6.3.1 Medical uses of RNA circuits

RNA circuits, and CRISPR-based RNA sensors in particular, could have interesting medical applications. Similar small molecule responsive gRNAs have already been shown to work in human cells (Liu et al., 2016), and RNA-sensing gRNAswitches could be used to construct 'smart drugs' responsive to the presence of specific RNAs. Although programmable RNA sensors exist for bacteria (Green et al., 2014), there are none available in eukaryotes. The simple and programmable mechanism of gRNAswitches, combined with the different output platforms based on the Cas9 protein, could form the basis of RNA-based 'smart drugs'. Programmable inputs could be the presence of viral RNAs, or if associated with small-molecule sensing gRNAs developed by Liu et al. (2016), combinations of RNAs and metabolites associated with cancer Xie et al. (e.g. 2011). Conditionally activated outputs could be associated with circuits that cause cell death (Nissim and Bar-Ziv, 2010), express immunostimulatory compounds, or downregulate immune evasion genes (Dannull et al., 2007), by taking advantage of the different Cas9 variants that cleave, upregulate or downregulate the target. gRNAswitches, alone or as part of RNA circuits, could improve RNA vaccines (that were proposed by Andries et al., 2015), and could be combined with other RNA based drugs that are being developed (Burnett and Rossi, 2012; Sahin et al., 2014). This would require reliable methods for *in-vivo* delivery of RNAs, for which various viral and non-viral means are being explored (Kauffman et al., 2015).

Another issue that needs to be addressed for RNA parts to be used in medical applications is that foreign RNA is inflammatory in human cells, due to its recognition by the antiviral immune response pathway described in the introduction of chapter 4 (section 4.1.1). To avoid inflammatory side effects, non-natural amino acids such as pseudouridine can be incorporated into the sequences (Karikó et al., 2008; Anderson et al., 2010a). This also has the advantage of enhancing RNA stability (Karikó et al., 2008). However, these modifications are known to have different thermodynamic properties than ordinary bases (Serra et al., 2004), but have not been studied in detail (Mathews and Turner, 2006) and are not incorporated in secondary structure prediction software. New biochemical studies of the thermodynamic properties of RNAs containing modified bases that have relevant medical uses could facilitate the creation of RNA circuits for medical uses. In the case of gRNAswitches with modified nucleotides, their ability to bind the Cas9 protein would also need to be tested.

6.4 Concluding remarks

The versatility and programmability of RNA has permitted the development of many RNA-based parts and circuits. This work contributes to the expansion of this catalog through the development of design rules for two new types of RNA parts: Circular riboregulators and RNA-sensing CRISPR gRNAswitches. PIE ribozymes are demonstrated to be a flexible method for generating circular RNAs, simply by pasting the cargo sequence between the ribozyme exons. This can generate active, circular riboregulators in *E. coli*, although in this context they are generally less active than their counterparts. They nonetheless interact with their target through base-pairing interactions and are generated in sufficient quantities to control transcription from cis-repressed messenger RNAs. A method for improving riboregulators through directed evolution based on continuous growth of T7 bacteriophage was also tested. Although this did not yield improved RNA parts, we showed that recombination could take place between a phage and a transduced phagemid, which could allow the use of

library based methods to increase the genetic diversity during continuous evolution. Finally, a new type of RNA sensor was developed, based on engineered CRISPR-Cas9 gRNAs. We show that gRNAs can be repressed by the addition of a cis-regulating element which blocks the gRNA targeting region through base pairing. Further addition of toehold and loop regions which hybridise to a second trigger RNA allow these switches to conditionally target the Cas9 protein to its DNA target upon RNA trigger addition. These new RNA parts could find applications such as control elements in RNA circuits, which could be used within *in vitro* circuits, in programmable bacterial or eukaryotic cells, or for producing RNA-based drugs. Future testing of synthetic circRNA regulators in different contexts might reveal applications where circularity brings a significant advantage, whereas refinement of the design rules of gRNAswitches, and testing with a larger number of mRNA trigger sites could permit the improvement of their sensitivity, which would make them an attractive choice as synthetic RNA sensors in eukaryotic cells.

A Appendix: Supplementary methods

A.0.1 Supplemented M9 minimal medium

Ingredient	Concentration
$CaCl_2$	100 µM
$MgSO_4$	2 mM
$FeSO_4$	10 µM
M9 Salts	1x
Casein amino acids	0.2% w/v
Glycerol	0.8% v/v
Thiamine	1 µg/mL
Uracil	20 µg/mL
Leucine	30 µg/mL
NaOH	Adjust to pH 7.4

Table A.1 – Composition of enriched M9 medium.

A.1 Generic goldengate cloning plasmids

A.1.1 Multiple cloning sites

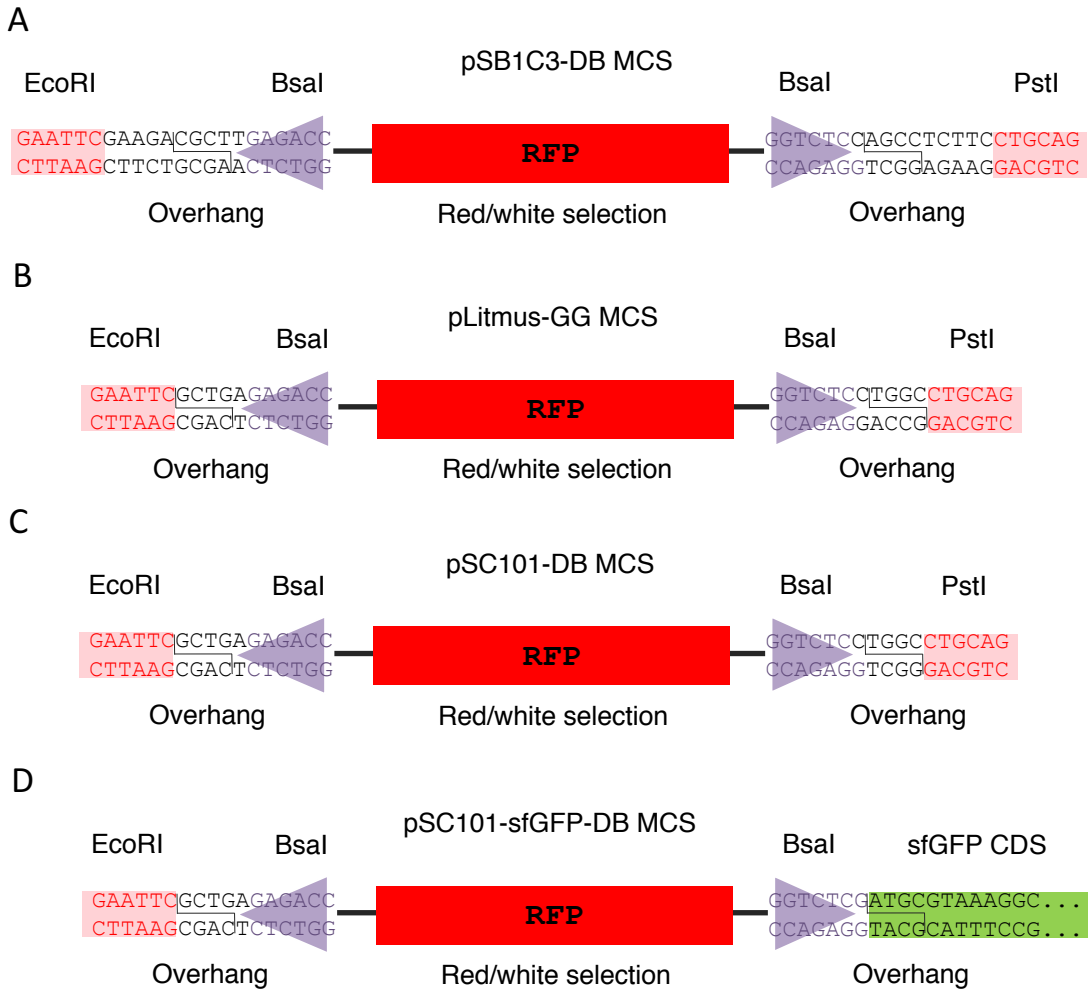


Figure A.1 – Multiple cloning sites used in generic biobrick compatible goldengate plasmids. **A, B, C:** Cloning sites for the pSB1C3, pLitmus-GG and pSC101-DB plasmids, respectively. EcoRI and PstI sites surround the divergent BsaI insertion sites. Parts can be inserted by goldengate, and the RFP cassette allows red/white selection. **D.** The pSC101-sfGFP-DB plasmid has a similar upstream site and RFP cassette, but the downstream BsaI site overhang is the start codon of sfGFP, allowing easy insertion of promoters or 5' UTR sequences.

A.1.2 Plasmid maps

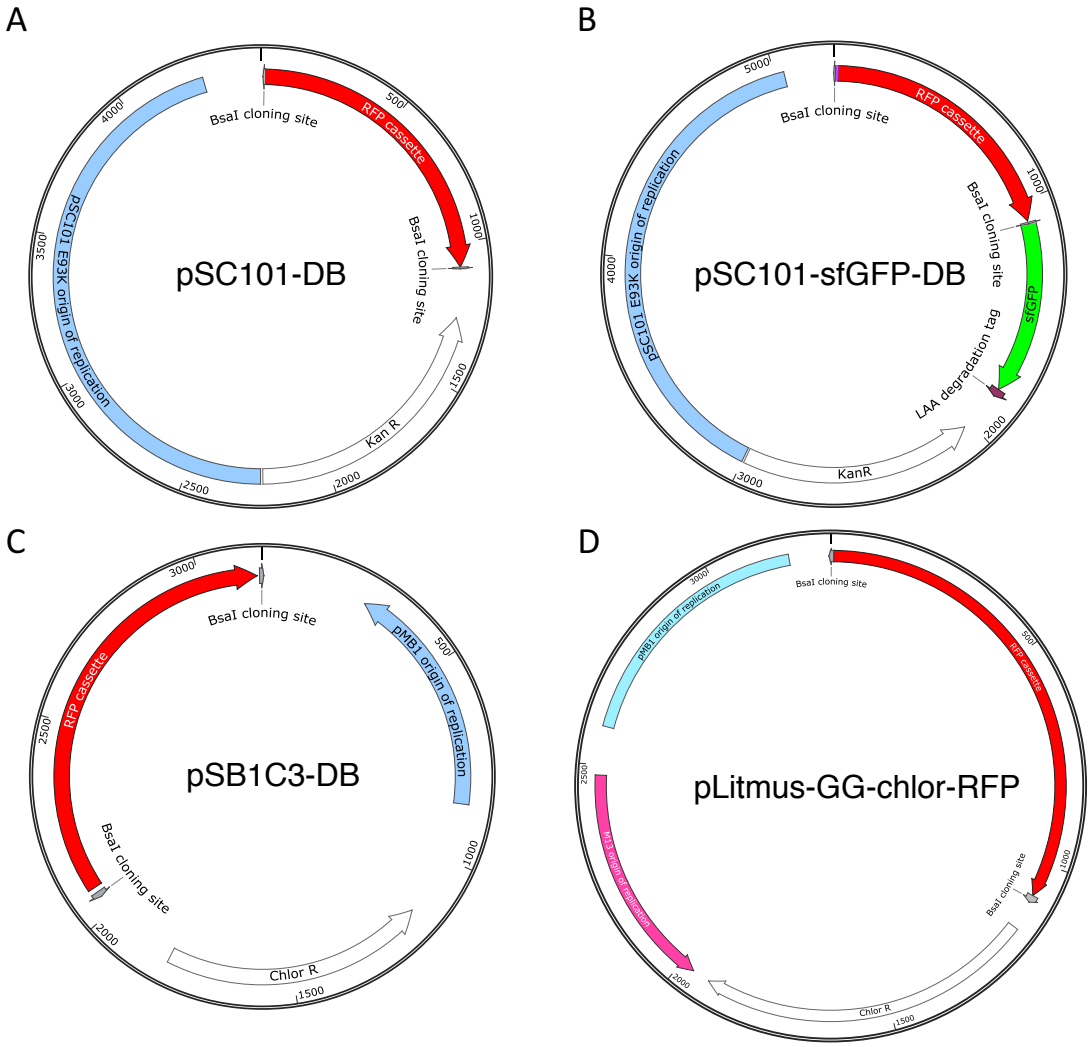


Figure A.2 – Plasmid maps of generic goldengate cloning plasmids. the RFP cassette is the J04450 biobrick part. pSC101-DB and pSC101-sfGFP-DB are derived from the pSB4K5 biobrick plasmid, pSB1C3-DB is derived from pSB1C3, and pLitmus-GG-chlor-RFP is derived from pLitmus28.

B Appendix to chapter 2

B.0.1 Sequences

Part	Sequence
PIE ribozyme (Cargo = NNNN)	TCTAGAGGTTCTACATAAATGCCTAACGACTATCCCTTTGGGGAGTAGGG TCAAGTGACTCGAAACGATAGACAACTTGCTTTAACAAGTTGGAGATATA GTCTGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATT AACGACCTTATCTGAACATAATGCTACCGTTTAATATTNNNNATGTTTTCTT GGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAA ACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTCAAT TGATCCTTAGCGAAAGCTAAGGATTTTTTTT
PIE ribozyme C873U	TCTAGAGGTTCTACATAAATGCCTAACGATTATCCCTTTGGGGAGTAGGG TCAAGTGACTCGAAACGATAGACAACTTGCTTTAACAAGTTGGAGATATA GTCTGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATT AACGACCTTATCTGAACATAATGCTACCGTTTAATATTNNNNATGTTTTCTT GGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAA ACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTCAAT TGATCCTTAGCGAAAGCTAAGGATTTTTTTT
circRAJ31 (insert)	CCTAGGGGATCAATATTAAGCTTTACCGTCGTCTCATCCGCAATTAAGTC CCTATCCGCTGTCACTGTATTGGTCTGCTAGC
circRAJ31Toe	CCTAGGGGATCAATATTAAGCTTTACCGACACCCCATCCGCAATTAAGTC CCTATCCGCTGTCACTGTATTGGTCTGCTAGC
Nusl3	GGATAGGGACTTAATTGCGGATGAGACGACGGTAAATTATTAATTAACTT AAGGAGGTAATAAATGACCGTCGTCTGGGCCACTGGGCGTCATACAATG ACTAGC
Nusl6	GGGACTTAATTGCGGATGAGACGACGGTAAGCGTTAATTAACTTAAGGAG GTAACGCATGCCGTCGTCTCAAGGCTTTAGCCGTCACTTTATGACTAGC
taRAJ31min	GGATCAATATTAAGCTTTACCGTCGTCTCATCCGCAATTAAGTCCCTATCC GCTGTCACTGTATTGGTCTATCCTTAGCGAAAGCTAAGGATTTTTTTT
circRAG1v1	CCTAGGGCAGGGATAAACGAGATAGATAAGATAAGATAGCCTAGG
circRAG1v2	CCTAGGGCCGGCAGGGATAAACGAGATAGATAAGATAAGATAGCGGCCC TAGG
linRAG1	GGGACTGACTATTCTGTGCAATAGTCAGTAAAGCAGGGATAAACGAGATA GATAAGATAAGATAGATCCTTAGCGAAAGCTAAGGATTTTTTTT
cisRAG1	GGGTCTTATCTTATCTATCTCGTTTATCCCTGCATACAGAAACAGAGGAGA TATGCAATGATAAACGAGAACCTGGCGGCAGCGCAAAAG
circRAG2v2	CCTAGGGCCGCACTAAACTAAACGACAATGTAATCAAACTCGGCCCTAGG
linRAG2	GGGACAGATCCACTGAGGCGTGGATCTGTGAACACTAAACTAAACGACAATGTAATC AAACTAACATCCTTAGCGAAAGCTAAGGATTTTTTTT
cisRAG2	GGGAGTTTGATTACATTGTCGTTTAGTTTAGTGATACATAAACAGAGGAGA TATCACATGACTAAACGAAACCTGGCGGCAGCGCAAAAG
circRAG5v1	GAAACTAACAGAAGCCAAATCAATTACATAC

Table B.1 – List of sequences used in chapter 2.

Part	Sequence
PIE ribozyme (Cargo = NNNN)	TCTAGAGGTTCTACATAAATGCCTAACGACTATCCCTTTGGGGAGTAGGG TCAAGTGACTCGAAACGATAGACAACTTGCTTTAACAAGTTGGAGATATA GTCTGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATT AACGACCTTATCTGAACATAATGCTACCGTTTAATATTNNNNATGTTTTCTT GGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAA ACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTCAAT TGATCCTTAGCGAAAGCTAAGGATTTTTTTT
PIE ribozyme C873U	TCTAGAGGTTCTACATAAATGCCTAACGATTATCCCTTTGGGGAGTAGGG TCAAGTGACTCGAAACGATAGACAACTTGCTTTAACAAGTTGGAGATATA GTCTGCTCTGCATGGTGACATGCAGCTGGATATAATTCCGGGGTAAGATT AACGACCTTATCTGAACATAATGCTACCGTTTAATATTNNNNATGTTTTCTT GGGTTAATTGAGGCCTGAGTATAAGGTGACTTATACTTGTAATCTATCTAA ACGGGGAACCTCTCTAGTAGACAATCCCGTGCTAAATTGTAGGACTCAAT TGATCCTTAGCGAAAGCTAAGGATTTTTTTT
circRAJ31 (insert)	CCTAGGGGATCAATATTAAGCTTTACCGTCGTCTCATCCGCAATTAAGTC CCTATCCGCTGTCACTGTATTGGTCTGCTAGC
circRAJ31Toe	CCTAGGGGATCAATATTAAGCTTTACCGACACCCCATCCGCAATTAAGTC CCTATCCGCTGTCACTGTATTGGTCTGCTAGC
Nusl3	GGATAGGGACTTAATTGCGGATGAGACGACGGTAAATTATTAATTAACTT AAGGAGGTAATAAATGACCGTCGTCTGGGCCACTGGGCGTCATACAATG ACTAGC
Nusl6	GGGACTTAATTGCGGATGAGACGACGGTAAGCGTTAATTAACTTAAGGAG GTAACGCATGCCGTCGTCTCAAGGCTTTAGCCGTCACTTTATGACTAGC
taRAJ31min	GGATCAATATTAAGCTTTACCGTCGTCTCATCCGCAATTAAGTCCCTATCC GCTGTCACTGTATTGGTCTATCCTTAGCGAAAGCTAAGGATTTTTTTT
circRAG1v1	CCTAGGGCAGGGATAAACGAGATAGATAAGATAAGATAGCCTAGG
circRAG1v2	CCTAGGGCCGGCAGGGATAAACGAGATAGATAAGATAAGATAGCGGCCC TAGG
linRAG1	GGGACTGACTATTCTGTGCAATAGTCAGTAAAGCAGGGATAAACGAGATA GATAAGATAAGATAGATCCTTAGCGAAAGCTAAGGATTTTTTTT
cisRAG1	GGGTCTTATCTTATCTATCTCGTTTATCCCTGCATACAGAAACAGAGGAGA TATGCAATGATAAACGAGAACCTGGCGGCAGCGCAAAAG
circRAG2v2	CCTAGGGCCGCACTAAACTAAACGACAATGTAATCAAACTCGGCCCTAGG
linRAG2	GGGACAGATCCACTGAGGCGTGGATCTGTGAACACTAAACTAAACGACAATGTAATC AAACTAACATCCTTAGCGAAAGCTAAGGATTTTTTTT
cisRAG2	GGGAGTTTGATTACATTGTCGTTTAGTTTAGTGATACATAAACAGAGGAGA TATCACATGACTAAACGAAACCTGGCGGCAGCGCAAAAG
circRAG5v1	GAAACTAACAGAAGCCAAATCAATTACATAC

Table B.2 – List of sequences used in chapter 2 (continued).

B.0.2 Primers used

Primer	Sequence
circRT-fw	GCAATTAAGTCCCTATCCGCTG
circRT-rv	GGATGAGACGACGGTAAAGC
tdC873U_sense	CTACATAAATGCCTAACGATTATCCCTTTGGGGAGTAGG
tdC873U_antisense	CCTACTCCCCAAAGGGATAATCGTTAGGCATTTATGTAG
wr_cr83	CACCGAATTCTCCCTATCAGTGATAGAGATTGACATCCCTATCAGTGATAGAGATACTGAGCACGGATCAATATTAAGCTTTACCGTCGTCTC
wr_cr84	GTCGATCCCTGCAGAAAAAAAATCCTTAGCTTTCGCTAAGGATAGACCAATACAGTGACAGCGGATAGACCAATACAGTGACAGCGGATAG
CircRT-Lin	GACATGCAGCTGGATATAATTCC
CircRaj_toe_sense	TGGTCTCTACCGACACCCCATCCGCAATTAAGTCCCTATCCGCTGTGAGACCTT
CircRaj_toe_asense	AAGGTCTCACAGCGGATAGGGACTTAATTGCGGATGGGGTGTCGGTAGAGACCA
GFP_AddLAA_fw	TGGTCTCTCAGCGTGATGATACTAGTAGCGGCCGC
GFP_AddLAA_Rv	AGGTCTCTTAGCTTTGTACAGTTCATCCATACCATGCG
LAA_anneal_Fw	TGGTCTCTGCTAGCGCAGCGAACGACGAAAATTACGCCCTTGCAGCAGAGACCT
LAA_anneal_Rv	AGGTCTCTGCTGCAAGGGCGTAATTTTCGTCGTTCGCTGCGCTAGCAGAGACCA
GG-CAT-Fw	TGGTCTCTATGGAGAAAAAAATCACTGGATATACCAC
GG-CAT-rv	AGGTCTCATTATTACGCCCCGCCCTGCC
cisR31-Fw	TGGTCTCTATAATACTAGTAGCGGCCGCTGC
cisR31-Rv	AGGTCTCACCATATTATGTTTCTCCTCTTTATGTCGTCTC
circRAG1v1S	TAGGGCAGGGATAAACGAGATAGATAAGATAAGATAGCCTAGG
circRAG1v1AS	ACATCCTAGGCTATCTTATCTTATCTATCTCGTTTATCCCTGC
circRAG4v1S	TAGGGACAGTAAACGGAGCTCTAATCCATAATCGCCTAGG
circRAG4v1AS	ACATCCTAGGCGATTATGGATTAGAGCTCCGTTTACTGTC
circRAG5v1S	TAGGGAAACTAACAGAAGCCAAATCAATTACATACCTAGG
circRAG5v1AS	ACATCCTAGGTATGTAATTGATTTGGCTTCTGTTAGTTTC
circRAG1v2S	TAGGGCCGGCAGGGATAAACGAGATAGATAAGATAAGATAGCGGCCCTAGG
circRAG1v2AS	ACATCCTAGGGCCGCTATCTTATCTTATCTATCTCGTTTATCCCTGCCGGC
circRAG2v2S	TAGGGCCGCACTAAACTAAACGACAATGTAATCAAACTCGGCCCTAGG
circRAG2v2AS	ACATCCTAGGGCCGAGTTTGATTACATTGTCGTTTAGTTTAGTGCGGC

Table B.3 – List of primers used in chapter 2.

C Appendix to chapter 3

C.1 Sequences

Primer	Sequence
pT7-RBS	TAATACGACTCACTATAGGGAGTCGTCAATAGAACACATTAAGGAGGTAGA GA
SL	GGCGCCGTCGTCAT
UTR1	GGATAGGGACTTAATTGCGGATGAGACGACGGTAAAGCGTATGCAGAACA GAGGAGAATACGCATGACCGTCGTCTATCGCTTCAGCGGTCAGAGTATGAC CGCAATTGCACCGGTAATTACGATC
UTR2	GGATAGGGACTTAATTGCGGATGAGACGACGGTAAAGCCTATGTGAAACAG AGGAGAATAGGCATGACCGTCGTCTATCGCCTTCGCGGTCACACTATGACC GCAATTGCACCGGTAATTACGATC
cmk_5'_30bp	ATGACCGCAATTGCACCGGTAATTACGATC

Table C.1 – List of sequences used in chapter 3.

C.2 Primers

Primer	Sequence
G5F_new	GGTCGCGTCAATGAAGCTATGG
G5R_new	AAAGAGAGCATCAAGGAACTCAAAG
CiE20	CCAGTGGTCTCCAACCCAGC
PR2	TTGGTCTCTGCCAGGTAATAGACAGTGTACCCAGC
PR1	TTGGTCTCTGCTGTGTTACTCCACGCGGTGC
pT7_anneal_sense	GTAATAATACGACTCACTATAGGGAGTCGTCAATAGAACACATTAAGGA GGTAGAGA
pT7_anneal_asense	TCATTCTCTACCTCCTTAATGTGTTCTATTGACGACTCCCTATAGTGAGT CGTATTA
SL_anneal_sense	GTAAGGCGCCGTCGTCAT
SL_anneal_asense	TCATATGACGACGGCGCC

Table C.2 – List of primers used in chapter 3.

C.3 Bioreactor construction

The bioreactors were built by Mr. Fabio Polesel using customised glass bottles and caps (see figure C.1. Detailed instructions for bioreactor construction are available on the Evoprog consortium website: http://evoprog.eu/ (Polesel et al., 2016).

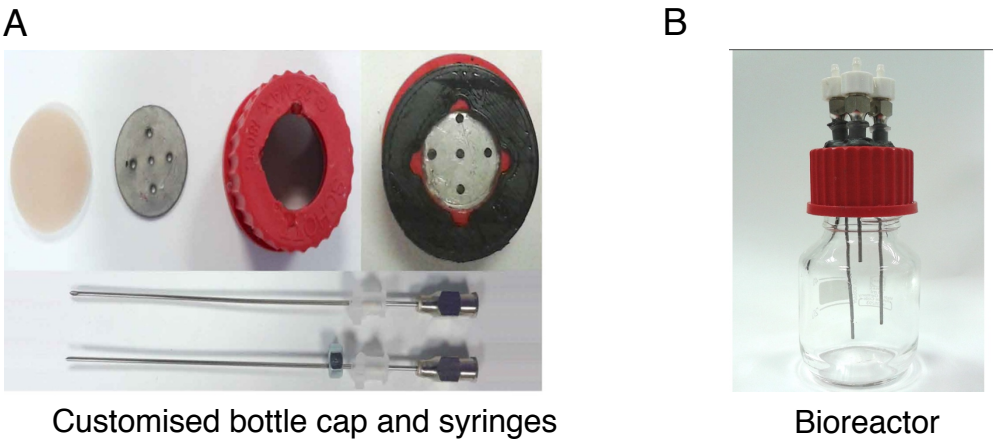


Figure C.1 – **A**. Caps and syringes used for input and output of media, cells and air. **B**. Assembled bottle.

D Supplementary methods: Chapter 4

D.1 Sequences used

Part	Sequence
J23119	TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC
HDV ribozyme	GGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATGC TTCGGCATGGCGAATGGGAC
Switch 3A	GCGACTTCATGCTGTAAAGGTTTTGCACGTTTCTTTGTCTGCCGTGCAAA ACCTTTCGCGGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTA GTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 7bp	TTTGCACGTTTCTTTGTCTGCCGTGCAAAACCTTTCGCGGTAGTTTTAGAG CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 8bp	GTGCAAAACCTTTCGCGGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 9bp	GTGCAAAACCTTTCGCGGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 10bp	GTGCAAAACCTTTCGCGGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 11bp	GTGCAAAACCTTTCGCGGTAGTTTTAGAGCTAGAAATAGCAAGTTAAAATA AGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 12bp	AAGGTTTTGCACGTTTCTTTGTCTGCCGTGCAAAACCTTTCGCGGTAGTTT TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAA AAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 13bp	AAAGGTTTTGCACGTTTCTTTGTCTGCCGTGCAAAACCTTTCGCGGTAGTT TTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG
Switch 3A No Toehold, 14bp	GAAAGGTTTTGCACGTTTCTTTGTCTGCCGTGCAAAACCTTTCGCGGTAG TTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT GAAAAAGTGGCACCGAGTCGGTGCTTTT
Switch 3A No Toehold, 15bp	CGAAAGGTTTTGCACGTTTCTTTGTCTGCCGTGCAAAACCTTTCGCGGTA GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT TGAAAAAGTGGCACCGAGTCGGTGCTTTT
Crisprzyme 1	GAAAGGCCCTTGGCAGGTGTCCTGGATTCCACCGTGTTGATTAAAGGTTTTAGAGCTA GAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAG TCGGTGCTTTTTTT
crisprzyme 2	TTTTGCACCTGATGAGCCTGGATACCAGCCGAAAGGCCCTTGGCAGTTA GACGAAACGGTGAAAGCCGTAGTGCAAAACCTTTCGCGGTAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAG TGGCACCGAGTCGGTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTG GCGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGGGAC
G12A crisprzyme	GACAAAACGGTGAAAGCCGTAGTGCAAAACCTTTCGCGGTAGTTTTAGAG CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGT GGCACCGAGTCGGTGCTTTT

Table D.1 – List of sequences used in chapter 4.

D.2 Primers used

Primer	Sequence
GR21	TTTTGGTCTCGGCTAGCATTATACCTAGGACTGAGCTAGC
GR22	TTTTGGTCTCTTAGCTTTGCACGTTTCTTTGTCTGCCG
GR23	TTTTGGTCTCTTAGCTTTTGCACGTTTCTTTGTCTGCCG
GR24	TTTTGGTCTCTTAGCGTTTTGCACGTTTCTTTGTCTGCCG
GR25	TTTTGGTCTCTTAGCGGTTTTGCACGTTTCTTTGTCTGCC
GR26	TTTTGGTCTCTTAGCAGGTTTTGCACGTTTCTTTGTCTGC
GR27	TTTTGGTCTCTTAGCAAGGTTTTGCACGTTTCTTTGTCTGC
GR28	TTTTGGTCTCTTAGCAAAGGTTTTGCACGTTTCTTTGTCTGC
GR29	TTTTGGTCTCTTAGCGAAAGGTTTTGCACGTTTCTTTGTCTGC
GR30	TTTTGGTCTCTTAGCCGAAAGGTTTTGCACGTTTCTTTGTCTGC
GR3A-2_rv	TATGGTCTCTGGTTTTGCACGTAATAGACGCCAGGGTGCAAAACCTTTCGC GGTAGTTTTAG
GR3A-2_fw	TATGGTCTCTAACCTTTCGTCTGGAACGTATCGCTAGCATTATACCTAGGAC TGAGCTAGC
GR3A-3_rv	TATGGTCTCTAACCTTTCGCGCAAAACTTTCAGCTAGCATTATACCTAGGAC TGAGCTAGC
GR3A-3_fw	TATGGTCTCTGGTTTTGCACGACGTCGCCAGGATGTGTGCAAAACCTTTCG CGGTAGTTTTAG
oJD46f	TAATACGACTCACTATAGGGTTTTGCACCTGATGAGCCTGGATACC
oJD46r	GTCCCATTCGCCATGCCGAAGC

Table D.2 – List of primers used in chapter 4.

E Appendix: NUPACK code

E.1 Design code for new RAJ31 activated UTRs

NUPACK design code for designing cis-repressed UTRs on a sfGFP gene, used to design Nusl3. The Nusl6 part is similar, but with a toehold shrank to 15 bases, as described in figure 2.13.

```

# NUPACK code for designing a cis-repressed UTR activated by circRAJ31
# William Rostain, U. of Warwick, 2016
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options
# material, sodium, magnesium, and dangles
#
material = rna # Use RNA
temperature[C] = 37.0    # optional units: C (default) or K
trials = 10 # The results will then be analysed manually
sodium[M] = 1.0    # optional units: M (default), mM, uM, nM, pM
dangles = some
#
# target structure using DU+ notation. See NUPACK website for info on notation.
structure cisrepressedUTR = U6U17D10(U3D6(U15)U3)U2D3(U4)D4(U5)U4
#
# sequence domains
#
# a = toehold and stem
```

```
# b = weak top stem
# c = loop region
# RBS = RBS
#
#Domains, they will be annotated on the NUPACK output
domain a = ggatagggacttaattgcggatgagacgacggtaaa
domain b = N6
domain c = N4
domain RBS = AACAGAGGAGA
domain d = ATAN3
domain ATG = ATG
domain e = N10
domain f = N12
domain g = N8
domain gfp = ATGactagc
#
# thread sequence domains onto target structures
#
cisrepressedUTR.seq = a b c RBS d ATG e f g gfp
#
# specify stop conditions for normalized ensemble defect
# default: 1.0 (percent) for each target structure
#
cisrepressedUTR.stop = 1.0

#
# prevent sequence patterns
#
prevent = AAAA, CCCC, GGGG, UUUU
prevent = KKKKKK, MMMMMM, RRRRRR, SSSSSS, WWWWWW, YYYYYY
```

E.2 Design code for mKate2 sensing gRNAswitches

Design code for gRNAswitches that use a full length RNA as an input.

```

# NUPACK code for designing a gRNAswitch, aka griboswitch
# Griboswitch initial design was conceived by the brilliant minds of Dr. J.
  Duncan and W. Rostain
# Code by William Rostain, U. of Warwick, 2016
################################################
```

```
#
# design material, temperature, and trials
# see NUPACK User Guide for valid options
# material, sodium, magnesium, and dangles
#
material = rna
temperature[C] = 37.0    # optional units: C (default) or K
trials = 10 # 10 sequences designed
sodium[M] = 1.0# optional units: M (default), mM, uM, nM, pM
dangles = some
#magnesium[M] = 0.05
##################################################
#
structure signal1 = U15U9U15 #This is the trigger RNA
structure griboRNA = U15 D13 (U15) U7U1D6(U1D4(U4)U3)U3D2(U5)U7D4(U4)U1D6(U3)U7 #
   This is the structure of the gRNAswitch alone
structure compl = D15(U9D15 + U13 ) U13 U7 U1D6(U1D4(U4)U3)U3D2(U5)U7D4(U4)U1D6(U
  3)U7 # This is the structure of the gRNAswitch in complex with the trigger
#
#
# The trigger is broken into three domains (5prime, spacer, 3prime)
domain sig1 = N15
domain sig2 = N9
domain sig3 = N15
#
# The switch is broken down into toehold, clamp (free), loop, OFFclamp (the 5
  prime part of targeting region that gets blocked), and Cas9 (which includes
  terminator)
domain toehold = N15 #The 15 bases toehold that is complementary to target
domain free = N13 #12 bases of the heteroduplexed loop
domain loop = N15 #15 bases of the loop, complementary to target
domain OFFclamp = N13 #12 bases that are occluded in the off state and become
  free. Normally at gRNA 5'
domain rest = N7 #8 free bases that are not in the loop, just before the cas9
  handle
domain free2 = N13
domain cas9 =
  GTTTTAGAGCTAGAAATAGCAAGnTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT

#
# Source can be any RNA
source RFP =
  atggcttcctccgaagacgttatcaaagagttcatgcgtttcaaagttcgtatggaaggttccgttaacggtcacgagttcga
```

```
  aatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccagaccgctaaactgaaagttaccaaaggtggtccgctgccgt

  tcgcttgggacatcctgtccccgcagtccagtacggttccaaagcttacgttaaacacccggctgacatcccggactacctga
   aactgtcc
#
#
window rfp_window = sig1 sig2 sig3
rfp_window.source = RFP
rfp_window.accessible[%] = [50,100]     # optional units: % (default) or frac.
  Input accessibility here [lower limit, upper limit]
#
#
#
strand griboRNAstrand = toehold free2 loop OFFclamp rest cas9
strand signal = sig1 sig2 sig3
griboRNA.seq = griboRNAstrand
signal1.seq = signal
compl.seq = signal griboRNAstrand
signal1.stop[%] = 25.0  # optional units: % or frac
## Illegal sequence, relaxed
prevent = AAAA, CCC
```

References

Abe, N., Abe, H., and Ito, Y. (2007). Dumbbell-shaped nanocircular RNAs for RNA interference. *Journal of the American Chemical Society*, 129(49):15108–15109.

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B. T., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., Severinov, K., Regev, A., Lander, E. S., Koonin, E. V., and Zhang, F. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (New York, N.Y.)*, 353(6299):aaf5573.

Andersen, J., Delihas, N., Ikenaka, K., Green, P. J., Pines, O., Ilercil, O., and Inouye, M. (1987). The isolation and characterization of RNA coded by the micF gene in Escherichia coli. *Nucl. Acids Res.*, 15(5):2089–2101.

Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjørn, S. P., Givskov, M., and Molin, S. (1998). New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria. *Applied and environmental microbiology (1998)*, 64(6):2240–2246.

Anderson, B. R., Muramatsu, H., Nallagatla, S. R., Bevilacqua, P. C., Sansing, L. H., Weissman, D., and Karik, K. (2010a). Incorporation of pseudouridine into mRNA enhances translation by diminishing PKR activation. *Nucleic Acids Research*, 38(17):5884–5892.

Anderson, J. C., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. a., Arkin, A. P., and Keasling, J. D. (2010b). BglBricks: A flexible standard for biological part assembly. *Journal of biological engineering*, 4(1):1.

Ando, H., Lemire, S., Pires, D. P., and Lu, T. K. (2015). Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Systems*, 1(3):187–196.

Andries, O., Kitada, T., Bodner, K., Sanders, N. N., and Weiss, R. (2015). Synthetic biology devices and circuits for RNA-based 'smart vaccines': a propositional review. *Expert review of vaccines*, 14(2):313–31.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology*, 2:2006.0008.

Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016). Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature*, 533(7601):58–63.

Badran, A. H. and Liu, D. R. (2015). Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nature communications*, 6:8425.

Banerjee, A. R., Jaeger, J. A., and Turner, D. H. (1993). Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry*, 32(1):153–163.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. a., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)*, 315(5819):1709–12.

Barrett, S. P. and Salzman, J. (2016). Circular RNAs: analysis, expression and potential functions. *Development*, 143(11):1838–1847.

Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999). Tertiary motifs in RNA structure and folding. *Angewandte Chemie - International Edition*, 38(16):2326–2343.

Been, M. D. (1996). Circular Ribozymes Generated in Escherichia coli Using Group I Self-splicing Permuted Intron-Exon Sequences. *Journal of Biological Chemistry*, 271(42):26081–26087.

Belfort, M., Chandry, P., and Pedersen-Lane, J. (1987). Genetic delineation of functional components of the group I intron in the phage T4 td gene. *Cold Spring Harb Symp Quant Biol.*, 52:181–192.

Berens, C. and Suess, B. (2015). Riboswitch engineering — making the all-important second and third steps. *Current Opinion in Biotechnology*, 31:10–15.

Bernick, D. L., Dennis, P. P., Lui, L. M., and Lowe, T. M. (2012). Diversity of antisense and other non-coding RNAs in archaea revealed by comparative small RNA sequencing in four Pyrobaculum species. *Frontiers in Microbiology*, 3(JUL):1–18.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., and Long, R. M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, 2(4):437–445.

Bessette, P. H., Rice, J. J., and Daugherty, P. S. (2004). Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Engineering, Design and Selection*, 17(10):731–739.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research*, 41(15):7429–7437.

Borujeni, A. E., Mishler, D. M., Wang, J., Huso, W., and Salis, H. M. (2015). Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Research*, page gkv1289.

Borujeni, A. E. and Salis (2011). Small RNA calculator. URL: https://salislab.net/software/SmallRNACalculator_ReverseEng. Last accessed: 20.01.2017.

Brion, P., Michel, F., Schroeder, R., and Westhof, E. (1999). Analysis of the cooperative thermal unfolding of the td intron of bacteriophage T4. *Nucleic acids research*, 27(12):2494–502.

Brown, T. (2010). *Introduction of DNA into Living Cells*. Wiley-Blackwell, Manchester, 6th editio edition.

Burnett, J. C. and Rossi, J. J. (2012). RNA-based therapeutics: Current progress and future prospects. *Chemistry and Biology*, 19(1):60–71.

Byun, J., Lan, N., Long, M., and Sullenger, B. A. (2003). Efficient and specific repair of sickle β-globin RNA by trans-splicing ribozymes. *RNA*, 9(10):1254–1263.

Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. *Nature reviews. Microbiology*, 12(5):381–90.

Carlson, J. C., Badran, A. H., Guggiana-Nilo, D. A., and Liu, D. R. (2014). Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol*, 10(3):216–222.

Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981). In vitro splicing of the ribosomal RNA precursor of tetrahymena: Involvement of a guanosine nucleotide in the excision of the intervening sequence. *Cell*, 27(3 PART 2):487–496.

Chandry, P. S. and Belfort, M. (1987). Activation of a cryptic 5' splice site in the upstream exon of the phage T4 td transcript: exon context, missplicing, and mRNA deletion in a fidelity mutant. *Genes & Development*, 1(9):1028–1037.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., Park, J., Blackburn, E. H., Weissman, J. S., Qi, L. S., and Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, 155(7):1479–1491.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. *Molecular Systems Biology*, 11(1):781.

Chen, L.-l. (2016a). The biogenesis and emerging roles of circular RNAs. *Nature reviews. Molecular cell biology*, 17(4):205–11.

Chen, L.-l. (2016b). The biogenesis and emerging roles of circular RNAs. *Nature reviews. Molecular cell biology*, 17(4):205–11.

Coelho, P. S., Brustad, E. M., Kannan, A., and Arnold, F. H. (2013). Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science (New York, N.Y.)*, 339(6117):307–10.

Coleman, J., Green, P. J., and Inouye, M. (1984). The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes. *Cell*, 37(2):429–436.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. a., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science (New York, N.Y.)*, 339(6121):819–23.

Coulon, A., Ferguson, M. L., De Turris, V., Palangat, M., Chow, C. C., and Larson, D. R. (2014). Kinetic competition during the transcription cycle results in stochastic RNA processing. *eLife*, 3:1–22.

Cruz-Toledo, J., McKeague, M., Zhang, X., Giamberardino, A., McConnell, E., Francis, T., DeRosa, M. C., and Dumontier, M. (2012). Aptamer base: A collaborative knowledge base to describe aptamers and SELEX experiments. *Database*, 2012:bas006–bas006.

Danan, M., Schwartz, S., Edelheit, S., and Sorek, R. (2012). Transcriptome-wide discovery of circular RNAs in Archaea. *Nucleic Acids Research*, 40(7):3131–3142.

Dannull, J., Lesher, D.-t., Holzknecht, R., Qi, W., Hanna, G., Seigler, H., Tyler, D. S., and Pruitt, S. K. (2007). Immunoproteasome down-modulation enhances the ability of dendritic cells to stimulate antitumor immunity Immunoproteasome down-modulation enhances the ability of dendritic cells to stimulate antitumor immunity. *Blood*, 110(13):4341–4350.

Darsonval, M., Msadek, T., Alexandre, H., and Grandvalet, C. (2015). The antisense RNA approach: a new application for in vivo stress response investigation in Oenococcus oeni. *Applied and Environmental Microbiology*, 82(October):AEM.02495–15.

de la Peña, M. and García-Robles, I. (2010). Ubiquitous presence of the hammerhead ribozyme motif along the tree of life. *RNA (New York, N.Y.)*, 16(10):1943–50.

de Silva, C. and Walter, N. G. (2009). Leakage and slow allostery limit performance of single drug-sensing aptazyme molecules based on the hammerhead ribozyme. *Rna*, 15(1):76–84.

de Vriend, H. (2006). *Early social reflections on the emerging field of synthetic biology*. Rathenau Institute, The Hague, 1rst edition.

Deana, A., Celesnik, H., and Belasco, J. G. (2008). The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature*, 451(7176):355–8.

Del Campo, C., Bartholomäus, A., Fedyunin, I., and Ignatova, Z. (2015). Secondary Structure across the Bacterial Transcriptome Reveals Versatile Roles in mRNA Regulation and Function. *PLoS Genetics*, 11(10):1–23.

Delebecque, C. J., Lindner, A. B., Silver, P. A., and Aldaye, F. A. (2011). Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science*, 333(6041):470–474.

Delihas, N. and Forst, S. (2001). MicF: an antisense RNA gene involved in response of Escherichia coli to global stress factors. *Journal of Molecular Biology*, 313(1):1–12.

Desai, R. P. and Papoutsakis, E. T. (1999). Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum. *Applied and Environmental Microbiology*, 65(3):936–945.

Desai, S. K. and Gallivan, J. P. (2004). Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *Journal of the American Chemical Society*, 126(41):13247–54.

Deutscher, M. P. (2006). Degradation of RNA in bacteria: Comparison of mRNA and stable RNA. *Nucleic Acids Research*, 34(2):659–666.

Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014). A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat Commun*, 5:5352.

Diener, T. O. (1989). Circular RNAs: relics of precellular evolution? *Proceedings of the National Academy of Sciences of the United States of America*, 86(23):9370–9374.

Dirks, R. M., Bois, J. S., Schaeffer, J. M., Winfree, E., and Pierce, N. a. (2007). Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Review*, 49(1):65–88.

Dirks, R. M., Lin, M., Winfree, E., and Pierce, N. a. (2004). Paradigms for computational nucleic acid design. *Nucleic Acids Research*, 32(4):1392–1403.

Dirks, R. M. and Pierce, N. a. (2003). Partition Function and Base-Pairing Probability Algorithms for Nucleic Acid Secondary Structure including Pseudoknots. *J Comput Chem*, 24:1664–1677.

du Lac, M., Scarpelli, A. H., Bates, D. G., and Leonard, J. N. (2016). Predicting the dynamics and heterogeneity of genomic DNA content within bacterial populations across variable growth regimes. *ACS Synthetic Biology*, page acssynbio.5b00217.

Dunn, J. J., Studier, F. W., and Gottesman, M. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology*, 166(4):477–535.

Durand, S., Gilet, L., Bessières, P., Nicolas, P., and Condon, C. (2012). Three essential ribonucleases-RNase Y, J1, and III-control the abundance of a majority of bacillus subtilis mRNAs. *PLoS Genetics*, 8(3).

Ellinger, T., Behnke, D., Knaus, R., Bujard, H., and Gralla, J. (1994). Context-dependent effects of upstream A-tracks. *Journal of molecular biology*, 239:466–475.

Ellington, A. D. and Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature*, 346(6287):818–822.

Endy, D. (2005). Foundations for engineering biology. *Nature*, 438(7067):449–453.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PloS one*, 3(11):e3647.

Englert, M. and Beier, H. (2005). Plant tRNA ligases are multifunctional enzymes that have diverged in sequence and substrate specificity from RNA ligases of other phylogenetic origins. *Nucleic Acids Research*, 33(1):388–399.

Enuka, Y., Lauriola, M., Feldman, M. E., Sas-Chen, A., Ulitsky, I., and Yarden, Y. (2015). Circular RNAs are long-lived and display only minimal early alterations in response to a growth factor. *Nucleic acids research*, (29):gkv1367–.

Espah Borujeni, A., Channarasappa, A. S., and Salis, H. M. (2014). Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research*, 42(4):2646–2659.

Esvelt, K. M., Carlson, J. C., and Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, 472(7344):499–503.

Esvelt, K. M., Mali, P., Braff, J. L., Moosburner, M., Yaung, S. J., and Church, G. M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nature methods*, 10(11):1116–21.

Farasat, I. and Salis, H. M. (2016). A Biophysical Model of CRISPR/Cas9 Activity for Rational Design of Genome Editing and Gene Regulation. *PLoS Computational Biology*, 12(1):1–33.

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet*, 9(2):102–114.

Fiskaa, T. and Birgisdottir, A. B. (2010a). RNA reprogramming and repair based on trans-splicing group I ribozymes. *New biotechnology*, 27(3):194–203.

Fiskaa, T. and Birgisdottir, Å. B. (2010b). RNA reprogramming and repair based on trans-splicing group I ribozymes. *New Biotechnology*, 27(3):194–203.

Flores, R., Gago-Zachert, S., Serra, P., Sanjuán, R., and Elena, S. F. (2014). Viroids: Survivors from the RNA World? *Annual review of microbiology*, (June):395–414.

Ford, E. and Ares, M. (1994). Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4. *Proceedings of the National Academy of Sciences of the United States of America*, 91(8):3117–21.

Fouts, D. E., True, H. L., and Celander, D. W. (1997). Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. *Nucleic acids research*, 25(22):4464–4473.

Franch, T., Petersen, M., Wagner, E. G. H., Jacobsen, J. P., and Gerdes, K. (1999). Antisense RNA regulation in prokaryotes: rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *Journal of molecular biology*, 294(5):1115–25.

Freier, S. M., Kierzek, R., Jaeger, J. a., Sugimoto, N., Caruthers, M. H., Neilson, T., and Turner, D. H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. *Proceedings of the National Academy of Sciences of the United States of America*, 83(24):9373–7.

Galloway Salvo, J. L., Coetzee, T., and Belfort, M. (1990). Deletion-tolerance and trans-splicing of the bacteriophage T4 td intron. Analysis of the P6-L6a region. *Journal of Molecular Biology*, 211(3):537–549.

Gao, Y. and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of Integrative Plant Biology*, 56(4):343–349.

Garcia, L. R. and Molineux, I. J. (1995). Rate of translocation of bacteriophage T7 DNA across the membranes of Escherichia coli. *Journal of Bacteriology*, 177(14):4066–4076.

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A., Lim, W. A., Weissman, J. S., and Qi, L. S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 154(2):442–451.

Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014). Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell*, 159(4):925–939.

Greener, A., Callahan, M., and Jerpseth, B. (1997). An efficient random mutagenesis technique using an E. coli mutator strain. *Molecular Biotechnology*, 7:189–195.

Guarnerio, J., Bezzi, M., Jeong, J. C., Paffenholz, S. V., Berry, K., Naldini, M. M., Lo-Coco, F., Tay, Y., Beck, A. H., and Pandolfi, P. P. (2016). Oncogenic Role of Fusion-circRNAs Derived from Cancer-Associated Chromosomal Translocations.

Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, 35(3 PART 2):849–857.

Guo, J. U., Agarwal, V., Guo, H., and Bartel, D. P. (2014). Expanded identification and characterization of mammalian circular RNAs. *Genome biology*, 15(7):409.

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., and Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, 495(7441):384–8.

Haurwitz, R. E., Jinek, M., Wiedenheft, B., Zhou, K., and Doudna, J. a. (2010). Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science (New York, N.Y.)*, 329(5997):1355–8.

He, L., Kierzek, R., Santalucia, J., Walter, a. E., and Turner, D. H. (1991). Nearest-Neighbor Parameters for G.U Mismatches - 5'gu3'/3'ug5' Is Destabilizing in the Contexts Cgug..Gugc, Ugua..Augu, and Aguu..Uugu but Stabilizing in Gguc..Cugg. *Biochemistry*, 30(46):11124–11132.

Hellsten, I. and Nerlich, B. (2011). Synthetic biology: building the language for a new science brick by metaphorical brick. *New Genetics and Society*, 30(March 2013):375–397.

Hemphill, J., Borchardt, E. K., Brown, K., Asokan, A., and Deiters, A. (2015). Optical control of CRISPR/Cas9 gene editing. *Journal of the American Chemical Society*, 137(17):5642–5645.

Hickey, D. R. and Turner, D. H. (1985). Effects of terminal mismatches on RNA stability: thermodynamics of duplex formation for ACCGGGp, ACCGGAp, and ACCGGCp. *Biochemistry*, 24(15):3987–91.

Hiszczyńska-Sawicka, E. and Kur, J. (1997). Effect of Escherichia coli IHF mutations on plasmid p15A copy number. *Plasmid*, 38(3):174–9.

Hochrein, L. M., Schwarzkopf, M., Shahgholi, M., Yin, P., and Pierce, N. A. (2013). Conditional dicer substrate formation via shape and sequence transduction with small conditional RNAs. *Journal of the American Chemical Society*, 135(46):17322–17330.

Hofacker, I. L. and Stadler, P. F. (2008). RNA Secondary Structures. In Lengauer, T., editor, *Bioinformatics-From Genomes to Therapies*, chapter 14, pages 439–489. Wiley-VCH Verlag GmbH, Weinheim, Germany, 1 edition.

Hoynes-O'Connor, A. and Moon, T. S. (2016). Development of design rules for reliable antisense RNA behavior in E. coli. *ACS Synthetic Biology*, page acssynbio.6b00036.

Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015). Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature methods*, 12(10):939–42.

Husimi, Y., Nishigaki, K., Kinoshita, Y., and Tanaka, T. (1982). Cellstat - a Continuous Culture System of a Bacteriophage for the Study of the Mutation Rate and the Selection Process At the Dna Level. *Review of Scientific Instruments*, 53(4):517–522.

Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. (1986). Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Research*, 14(9):3627–3640.

Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R., and Collins, J. J. (2004). Engineered riboregulators enable post-transcriptional control of gene expression. *Nature biotechnology*, 22(7):841–7.

Jaschke, P. R., Lieberman, E. K., Rodriguez, J., Sierra, A., and Endy, D. (2012). A fully decompressed synthetic bacteriophage Phi-X174 genome assembled and archived in yeast. *Virology*, 434(2):278–284.

Jeck, W. R. and Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nature biotechnology*, 32(5):453–61.

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F., and Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA (New York, N.Y.)*, 19(2):141–57.

Jensen, S. and Thomsen, a. R. (2012). Sensing of RNA Viruses: a Review of Innate Immune Receptors Involved in Recognizing RNA Virus Invasion. *Journal of Virology*, 86(January):2900–2910.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. A. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. *Science*, 348(6242):1477–1481.

Jiang, W., Samai, P., and Marraffini, L. A. (2016). Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell*, 164(4):710–721.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, 337(6096):816–21.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., Kaplan, M., Iavarone, A. T., Charpentier, E., Nogales, E., and Doudna, J. A. (2014). Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science (New York, N.Y.)*, 343(6176):1247997.

Jones, J. T., Lee, S. W., and Sullenger, B. A. (1996). Tagging ribozyme reaction sites to follow trans-splicing in mammalian cells. *Nature medicine*, 2(6):643–8.

Karikó, K., Muramatsu, H., Welsh, F. a., Ludwig, J., Kato, H., Akira, S., and Weissman, D. (2008). Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Molecular therapy : the journal of the American Society of Gene Therapy*, 16(11):1833–1840.

Kato, H., Fujisawa, H., and Minagawa, T. (1986). Subunit arrangement of the tail fiber of bacteriophage T3. *Virology*, 153(1):80–86.

Kauffman, K. J., Webber, M. J., and Anderson, D. G. (2015). Materials for Non-viral intracellular delivery of messenger RNA therapeutics. *Journal of controlled release : official journal of the Controlled Release Society*.

Kay, B. K., Adey, N. B., Yun-Sheng, H., Manfredi, J. P., Mataragnon, A. H., and Fowlkes, D. M. (1993). An M13 phage library displaying random 38-amino-acid peptides as a source of novel sequences with affinity to selected targets. *Gene*, 128(1):59–65.

Keiler, K. C., Waller, P. R., and Sauer, R. T. (1996). Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science (New York, N.Y.)*, 271(5251):990–3.

Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of biological engineering*, 3:4.

Kiani, S., Chavez, A., Tuttle, M., Hall, R. N., Chari, R., Ter-Ovanesyan, D., Qian, J., Pruitt, B. W., Beal, J., Vora, S., Buchthal, J., Kowal, E. J. K., Ebrahimkhani, M. R., Collins, J. J., Weiss, R., and Church, G. (2015). Cas9 gRNA engineering for genome editing, activation and repression. *Nature methods*, (September):1–6.

Kiro, R., Shitrit, D., and Qimron, U. (2014). Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA biology*, 11(1):42–4.

Kitney, R. and Freemont, P. (2012). Synthetic biology - The state of play. *FEBS Letters*, 586(15):2029–2036.

Kittleson, J. T., Wu, G. C., and Anderson, J. C. (2012). Successes and failures in modular genetic engineering. *Current Opinion in Chemical Biology*, 16(3-4):329–336.

Klauser, B. and Hartig, J. S. (2013). An engineered small RNA-mediated genetic switch based on a ribozyme expression platform. *Nucleic acids research*, pages 1–11.

Kobori, S. and Yokobayashi, Y. (2016). High-Throughput Mutational Analysis of a Twister Ribozyme. *Angewandte Chemie - International Edition*, 55(35):10354–10357.

Köhler, U., Ayre, B. G., Goodman, H. M., and Haseloff, J. (1999). Trans -splicing Ribozymes for Targeted Gene Delivery. *Journal of molecular biology*, 285(5):1935–1950.

Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., Platt, R. J., Scott, D. a., Church, G. M., and Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, 500(7463):472–6.

Kosuri, S., Goodman, D. B., Cambray, G., Mutalik, V. K., Gao, Y., Arkin, A. P., Endy, D., and Church, G. M. (2013). Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 110(34):14024–9.

Kushwaha, M., Rostain, W., Prakash, S., Duncan, J. N., and Jaramillo, A. (2016). Using RNA as molecular code for programming cellular function. *ACS Synthetic Biology*.

Labrou, N. E. (2010). Random mutagenesis methods for in vitro directed enzyme evolution. *Current protein & peptide science*, 11(1):91–100.

Laing, L. and Draper, D. (1994). Thermodynamics of RNA Folding in a Conserved Ribosomal RNA Domain.

Lan, N., Howrey, R. P., Lee, S.-W., Smith, C. A., and Sullenger, B. A. (1998). Ribozyme-Mediated Repair of Sickle β-Globin mRNAs in Erythrocyte Precursors. *Science*, 280(5369):1593–1596.

Lange, S., Katayama, Y., Schmid, M., Burkacky, O., Brauchle, C., Lamb D.C., D. C., and Jansen, R. P. (2008). Simultaneous transport of different localized mRNA species revealed by live-cell imaging. *Traffic*, 9(8):1256–1267.

Lasda, E. and Parker, R. (2014). Circular RNAs: diversity of form and function. *RNA (New York, N.Y.)*, 20(12):1829–42.

Lee, Y. J., Hoynes-O'Connor, A., Leong, M. C., and Moon, T. S. (2016). Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Research*, 44(5):2462–2473.

Lim, F. and Peabody, D. S. (2002). RNA recognition site of PP7 coat protein. *Nucleic acids research*, 30(19):4138–44.

Liu, Y., Zhan, Y., Chen, Z., He, A., Li, J., Wu, H., Liu, L., Zhuang, C., Lin, J., Guo, X., Zhang, Q., Huang, W., and Cai, Z. (2016). Directing cellular information flow via CRISPR signal conductors. *Nature Methods*, (September).

Lorenz, R., Bernhart, S. H., Höner Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB*, 6:26.

Lu, T. K. T., Koeris, M. S., Chevalier, B. S., Holder, J. W., McKenzie, G. J., and Brownell, D. R. (2013). Recombinant phage and methods.

Lutz, R. and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research*, 25(6):1203–10.

Lutz, S. (2011). Beyond directed evolution - semi-rational protein engineering and design. *Current opinion in biotechnology*, 21(6):734–743.

Mackie, G. A. (2000). Stabilization of circular rpsT mRNA demonstrates the 5'-end dependence of RNase E action in vivo. *The Journal of biological chemistry*, 275(33):25069–72.

Macnaughton, T. B., Shi, S. T., Modahl, L. E., and Lai, M. M. C. (2002). Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. *Journal of virology*, 76(8):3920–7.

Majorek, K. A., Dunin-Horkawicz, S., Steczkiewicz, K., Muszewska, A., Nowotny, M., Ginalski, K., and Bujnicki, J. M. (2014). The RNase H-like superfamily: New members, comparative structural analysis and evolutionary classification. *Nucleic Acids Research*, 42(7):4160–4179.

Mathews, D. H. (2004). Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *Rna*, 10:1178–1190.

Mathews, D. H., Banerjee, A. R., Luan, D. D., Eickbush, T. H., and Turner, D. H. (1997). Secondary structure model of the RNA recognized by the reverse transcriptase from the R2 retrotransposable element. *RNA (New York, N.Y.)*, 3(1):1–16.

Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004). Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19):7287–92.

Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology*, 288(5):911–40.

Mathews, D. H. and Turner, D. H. (2006). Prediction of RNA secondary structure by free energy minimization. *Current Opinion in Structural Biology*, 16(3):270–278.

McAllister, W. T. and Barrett, C. L. (1977). Roles of the early genes of bacteriophage T7 in shutoff of host macromolecular synthesis.McAllister, W.T. & Barrett, C.L., 1977. Roles of the early genes of bacteriophage T7 in shutoff of host macromolecular synthesis. Journal of virology, 23(3), pp.543–5. *Journal of virology*, 23(3):543–53.

McCafferty, J., Griffiths, a. D., Winter, G., and Chiswell, D. J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains.

McCaskill, J. S. (1990). The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29(6-7):1105–1119.

Meister, G. and Tuschi, T. (2004). Mechanisms of gene silencing by double-stranded RNA. *Nature*, 431(September):343–349.

Meluzzi, D., Olson, K. E., Dolan, G. F., Arya, G., and Müller, U. F. (2012). Computational prediction of efficient splice sites for trans-splicing ribozymes. *RNA (New York, N.Y.)*, 18(3):590–602.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., and Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, 495(7441):333–8.

Moak, M. and Molineux, I. J. (2000). Role of the Gp16 lytic transglycosylase motif in bacteriophage T7 virions at the initiation of infection. *Molecular Microbiology*, 37(2):345–355.

Molina-Sánchez, M. D., Martinez-Abarca, F., and Toro, N. (2006). Excision of the Sinorhizobium meliloti group II intron RmInt1 as circles in vivo. *Journal of Biological Chemistry*, 281(39):28737–28744.

Molineux, I. J. (2006). *The T7 group*. Oxford Uni Press, 2nd editio edition.

Moons, P., Faster, D., and Aertsen, A. (2013). Lysogenic conversion and phage resistance development in phage exposed Escherichia coli biofilms. *Viruses*, 5(1):150–161.

Morfeldt, E., Taylor, D., von Gabain, a., and Arvidson, S. (1995). Activation of alpha-toxin translation in Staphylococcus aureus by the trans-encoded antisense RNA, RNAIII. *The EMBO journal*, 14(18):4569–4577.

Morrison, T. G. and Malamy, M. H. (1971). T7 translational control mechanisms and their inhibiton by F factors. *Nat New Biol*, 231(19):37–41.

Müller, K. M. and Arndt, K. M. (2012). Standardization in synthetic biology. *Methods in Molecular Biology*, 813:23–43.

Na, D., Yoo, S. M., Chung, H., Park, H., Park, J. H., and Lee, S. Y. (2013). Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs. *Nature biotechnology*, 31(2):170–4.

Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., and Breaker, R. R. (2002). Genetic control by a metabolite binding mRNA. *Chemistry and Biology*, 9(9):1043–1049.

Nakashima, N., Tamura, T., and Good, L. (2006). Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli. *Nucleic Acids Research*, 34(20):1–10.

Nawtaisong, P., Fraser, M. E., Carter, J. R., and Fraser Jr., M. J. (2015). Trans-splicing group I intron targeting hepatitis C virus IRES mediates cell death upon viral infection in Huh7.5 cells. *Virology*, 481:223–234.

Nechaev, S. and Severinov, K. (1999). Inhibition of Escherichia coli RNA polymerase by bacteriophage T7 gene 2 protein. *Journal of molecular biology*, 289(4):815–826.

Nielsen, H. and Johansen, S. D. (2009). Group I introns Moving in new directions. *Rna Biology*, 6(4):375–383.

Nissim, L. and Bar-Ziv, R. H. (2010). A tunable dual-promoter integrator for targeting of cancer cells. *Molecular systems biology*, 6(444):444.

Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P., and Lu, T. K. (2014). Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Molecular Cell*, 54(4):698–710.

Ogawa, A. and Maeda, M. (2007). Aptazyme-based riboswitches as label-free and detector-free sensors for cofactors. *Bioorganic and Medicinal Chemistry Letters*, 17(11):3156–3160.

Olson, K. E. and Müller, U. F. (2012). An in vivo selection method to optimize trans-splicing ribozymes. *RNA (New York, N.Y.)*, 18(3):581–9.

Orencia, M. C., Yoon, J. S., Ness, J. E., Stemmer, W. P., and Stevens, R. C. (2001). Predicting the emergence of antibiotic resistance by directed evolution and structural analysis. *Nature structural biology*, 8(3):238–242.

Ostafe, R., Prodanovic, R., Lloyd Ung, W., Weitz, D. a., and Fischer, R. (2014). A high-throughput cellulase screening system based on droplet microfluidics. *Biomicrofluidics*, 8(4):041102.

Packer, M. S. and Liu, D. R. (2015). Methods for the directed evolution of proteins. *Nature reviews. Genetics*, 16(7):379–394.

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., Daleykeyser, A., Yin, P., and Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell*, 159(4):940–954.

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor,

D. H., Gehrke, L., and Collins, J. J. (2016). Rapid, Low-Cost Detection of Zika Virus Using Programmable Biomolecular Components. *Cell*, 165(5):1255–1266.

Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, 24(1):79–88.

Perriman, R. and Ares, M. (1998). Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA (New York, N.Y.)*, 4(9):1047–1054.

Peterson, J. and Phillips, G. J. (2008). New pSC101-derivative cloning vectors with elevated copy numbers. *Plasmid*, 59(3):193–201.

Petkovic, S. and Müller, S. (2015). RNA circularization strategies in vivo and in vitro. *Nucleic Acids Research*, 43(4):2454–2465.

Polesel, F., Rodrigez, R., and Jaramillo, A. (2016). Evoprog constortium open source hardware and software dissemination. http://Evoprog.eu. Accessed on 18.12.2016.

Puttaraju, M. and Been, M. D. (1992). Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons. *Nucleic acids research*, 20(20):5357–5364.

Qi, L. S., Larson, M. H., Gilbert, L. a., Doudna, J. a., Weissman, J. S., Arkin, A. P., and Lim, W. a. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell*, 152(5):1173–83.

Qimron, U., Marintcheva, B., Tabor, S., and Richardson, C. C. (2006). Genomewide screens for Escherichia coli genes affecting growth of T7 bacteriophage. *Proceedings of the National Academy of Sciences of the United States of America*, 103(50):19039–44.

Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014). Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences*, 111(50):201409523.

Ran, F. A., Hsu, P. D., Lin, C.-y., Gootenberg, J. S., Trevino, A., Scott, D. A., Inoue, A., Matoba, S., Zhang, Y., and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154(6):1380–1389.

Repoila, F. and Darfeuille, F. (2009). Small regulatory non-coding RNAs in bacteria: physiology and mechanistic aspects. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 101(2):117–31.

Rodrigo, G., Landrain, T. E., and Jaramillo, A. (2012). De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, 109(38):15271–6.

Rontó, G., Agamalyan, M. M., Drabkin, G. M., Feigin, L. A., and Lvov, Y. M. (1983). Structure of bacteriophage T7. Small-angle X-ray and neutron scattering study. *Biophysical journal*, 43(3):309–14.

Rostain, W., Landrain, T., Rodrigo, G., and Jaramillo, A. (2015). Regulatory RNA Design Through Evolutionary Computation and Strand Displacement. In Marchisio, M. A., editor, *Computational Methods in Synthetic Biology*, volume 1244 of *Methods in Molecular Biology*, pages 63–78. Springer New York.

Ruffner, D. E., Stormo, G. D., and Uhlenbeck, O. C. (1990). Sequence requirements of the hammerhead RNA self-cleavage reaction. *Biochemistry*, 29(47):10695–702.

Sahin, U., Karikó, K., and Türeci, Ö. (2014). mRNA-based therapeutics–developing a new class of drugs. *Nature reviews. Drug discovery*, 13(10):759–80.

Sakai, Y., Abe, K., Nakashima, S., Yoshida, W., Ferri, S., Sode, K., and Ikebukuro, K. (2014). Improving the gene-regulation ability of small RNAs by scaffold engineering in Escherichia coli. *ACS synthetic biology*, 3(3):152–62.

Salis, H. M., Mirsky, E. a., and Voigt, C. a. (2009). Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology*, 27(10):946–50.

Selinger, D. W., Saxena, R. M., Cheung, K. J., Church, G. M., and Rosenow, C. (2003). Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. *Genome Research*, 13(2):216–223.

Serganov, A. and Nudler, E. (2013). A decade of riboswitches. *Cell*, 152(1-2):17–24.

Serganov, A. and Patel, D. J. (2007). Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat. Rev. Genet.*, 8(10):776–790.

Serra, M. J., Axenson, T. J., and Turner, D. H. (1994). A model for the stabilities of RNA hairpins based on a study of the sequence dependence of stability for hairpins of six nucleotides. *Biochemistry*, 33(47):14289–14296.

Serra, M. J., Smolter, P. E., and Westhof, E. (2004). Pronounced instability of tandem IU base pairs in RNA. *Nucleic Acids Research*, 32(5):1824–1828.

Serra, M. J. and Turner, D. H. (1995). Predicting thermodynamic properties of RNA. In in Enzymology, B. T. M., editor, *Energetics of Biological Macromolecules*, volume Volume 259, pages 242–261. Academic Press.

Shatkin, A. J. (1976). Capping of eucaryotic mRNAs. *Cell*, 9(4 (part 2)):645–653.

Shen, S., Rodrigo, G., Prakash, S., Majer, E., Landrain, T. E., Kirov, B., Daròs, J.-A., and Jaramillo, A. (2015). Dynamic signal processing by ribozyme-mediated RNA circuits to control gene expression. *Nucleic acids research*, 43(10):5158–5170.

Shetty, R. P., Endy, D., and Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering*, 2(1):5.

Söte, S., Kleine, S., Schlicke, M., and Brakmann, S. (2011). Directed Evolution of an Error-Prone T7 DNA Polymerase that Attenuates Viral Replication. *ChemBioChem*, 12(10):1551–1558.

Soukup, G. a. and Breaker, R. R. (1999). Engineering precision RNA molecular switches. *Proceedings of the National Academy of Sciences*, 96(7):3584–3589.

Springman, R., Keller, T., Molineux, I. J., and Bull, J. J. (2010). Evolution at a high imposed mutation rate: Adaptation obscures the load in phage T7. *Genetics*, 184(1):221–232.

Steinberg, N. (2012). Flow Cytometry GUI for Matlab: https://fr.mathworks.com/matlabcentral/fileexchange/38080-flow-cytometry-gui-for-matlab. Accessed on 03.01.2017.

Stokke, C., Flåtten, I., and Skarstad, K. (2012). An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS ONE*, 7(2).

Studier, F. W. (1981). Identification and mapping of five new genes in bacteriophage T7. *Journal of Molecular Biology*, 153:493–502.

Sugimoto, N., Kierzek, R., and Turner, D. H. (1987). Sequence dependence for the energetics of terminal mismatches in ribooligonucleotides. *Biochemistry*, 26(14):4559–4562.

Šulc, P., Ouldridge, T. E., Romano, F., Doye, J. P. K., and Louis, A. A. (2015). Modelling toehold-mediated RNA strand displacement. *Biophysical Journal*, 108(5):1238–1247.

Sullenger, B. a. and Cech, T. R. (1994). Ribozyme-mediated repair of defective mRNA by targeted, trans-splicing.

Sun, Z. Z., Hayes, C. A., Shin, J., Caschera, F., Murray, R. M., and Noireaux, V. (2013). Protocols for implementing an Escherichia coli based TX-TL cell-free expression system for synthetic biology. *Journal of visualized experiments : JoVE*, (79):e50762.

Swe, P. M., Copp, J. N., Green, L. K., Guise, C. P., Mowday, A. M., Smaill, J. B., Patterson, A. V., and Ackerley, D. F. (2012). Targeted mutagenesis of the Vibrio fischeri flavin reductase FRase i to improve activation of the anticancer prodrug CB1954. *Biochemical Pharmacology*, 84(6):775–783.

Swers, J. S., Kellogg, B. A., and Wittrup, K. D. (2004). Shuffled antibody libraries created by in vivo homologous recombination and yeast surface display. *Nucleic acids research*, 32(3):e36.

Takahashi, M. K., Watters, K. E., Gasper, P. M., Abbott, T. R., Carlson, P. D., Chen, A. A., and Lucks, J. B. (2016). Using in-cell SHAPE-Seq and simulations to probe structure–function design principles of RNA transcriptional regulators. *RNA (New York, NY)*, pages 1–14.

Tee, K. L. and Wong, T. S. (2013). Polishing the craft of genetic diversity creation in directed evolution. *Biotechnology Advances*, 31(8):1707–1721.

Tenaillon, O., Barrick, J. E., Ribeck, N., Deatherage, D. E., Blanchard, J. L., Dasgupta, A., Wu, G. C., Schneider, D., and Lenski, R. E. (2016). Tempo and mode of genome evolution in a 50,000 generation experiment. *Nature*, 536(7615):1–21.

Thyme, S. B., Akhmetova, L., Montague, T. G., Valen, E., and Schier, A. F. (2016). Internal guide RNA interactions interfere with Cas9-mediated cleavage. *Nature Communications*, 7:11750.

Travisano, M., Vasi, F. K., and Lenski, R. E. (1995). Long-Term Experimental Evolution in Escherichia coli. III. Variation Among Replicate Populations in Correlated Responses to Novel Environments.

Tridgett, M. (2015). *In Vivo Recombineering of Bacteriophage T7 Imparts Host- Independence and High Fitness Cost.* Masters thesis, University of Warwick.

Tsai, S. Q., Wyvekens, N., Khayter, C., Foden, J. A., Thapar, V., Reyon, D., Goodwin, M. J., Aryee, M. J., and Joung, J. K. (2014). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature Biotechnology*, 32(6):569–576.

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science (New York, N.Y.)*, 249(4968):505–510.

Turner, D. H. and Mathews, D. H. (2010). NNDB: The nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Research*, 38(SUPPL.1):2009–2011.

Umekage, S. and Kikuchi, Y. (2009). In vitro and in vivo production and purification of circular RNA aptamer. *Journal of biotechnology*, 139(4):265–72.

Umekage, S., Uehara, T., Fujita, Y., Suzuki, H., and Kikuchi, Y. (2012). In Vivo Circular RNA Expression by the Permuted Intron-Exon Method. *Innovations in Biotechnology*, pages 75–90.

Updegrove, T. B., Zhang, A., and Storz, G. (2016). Hfq: The flexible RNA matchmaker. *Current Opinion in Microbiology*, 30:133–138.

Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G., and Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *Journal of molecular biology*, 270(5):724–738.

Varani, G. and McClain, W. H. (2000). The G·U wobble base pair. *EMBO reports*, 1(1):18–23.

Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–8.

Wang, Y.-h., Wei, K. Y., and Smolke, C. D. (2013). Systems biology: advancing the design of diverse genetic systems. *Annu. Rev. Chem. Biomol. Eng.*, (1):69–102.

Watanabe, T. and Sullenger, B. A. (2000). Induction of wild-type p53 activity in human cancer cells by ribozymes that repair mutant p53 transcripts. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8490–4.

Watters, K. E., Abbott, T. R., and Lucks, J. B. (2016). Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Research*, 44(2):e12.

Wieland, M., Benz, A., Klauser, B., and Hartig, J. S. (2009). Artificial ribozyme switches containing natural riboswiteh aptamer domains. *Angewandte Chemie - International Edition*, 48(15):2715–2718.

Win, M. N. and Smolke, C. D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences of the United States of America*, 104(36):14283–8.

Wolfe, B. R. and Pierce, N. A. (2014). Sequence Design for a Test Tube of Interacting Nucleic Acid Strands. *ACS Synthetic Biology*, page 141020092749006.

Wong, T. S., Zhurina, D., and Schwaneberg, U. (2006). The diversity challenge in directed protein evolution. *Combinatorial chemistry & high throughput screening*, 9(4):271–88.

Woodson, S. A. (2015). RNA folding retrospective: lessons from ribozymes big and small. *RNA (New York, NY)*, 21(4):502–503.

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences of the United States of America*, 112(11):3570–3575.

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., and Benenson, Y. (2011). Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science*, 333(6047):1307–1311.

Yang, Y., Lin, Y., Li, L., Linhardt, R. J., and Yan, Y. (2015). Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metabolic Engineering*, 29:217–226.

Ye, C. Y., Chen, L., Liu, C., Zhu, Q. H., and Fan, L. (2015). Widespread noncoding circular RNAs in plants. *New Phytologist*, 208(1):88–95.

Yen, L., Jennifer, S., Lee, J.-S., Gray, J. T., Magnier, M., Baba, T., D'Amato, R. J., and Mulligan, R. C. (2004). Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature*, 431(September):471–476.

Yoo, S. M., Na, D., and Lee, S. Y. (2013). Design and use of synthetic regulatory small RNAs to control gene expression in Escherichia coli. *Nature protocols*, 8(9):1694–707.

Young, E. M., Tong, A., Bui, H., Spofford, C., and Alper, H. S. (2014). Rewiring yeast sugar transporter preference through modifying a conserved protein motif. *Proceedings of the National Academy of Sciences of the United States of America*, 111(1):131–6.

Zadeh, J. N., Wolfe, B. R., and Pierce, N. A. (2011). Nucleic acid sequence design via efficient ensemble defect optimization. *Journal of computational chemistry*, 32(3):439–52.

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., and Lim, W. A. (2015). Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell*, 160(1):339–350.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., Volz, S. E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., and Zhang, F. (2015a). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3):759–771.

Zetsche, B., Volz, S. E., and Zhang, F. (2015b). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology*, 33(2):139–142.

Zhang, X.-O., Dong, R., Zhang, Y., Zhang, J.-L., Luo, Z., Zhang, J., Chen, L.-L., and Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome research*, 26(9):1277–87.

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. *Science (New York, N.Y.)*, 244(4900):48–52.

Zuker, M. and Stiegler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9(11981):133–148.